Universiteit
Leiden
The Netherlands

Determinants of genome editing outcomes: the impact of target and donor DNA structures

Chen, X.

Citation
Chen, X. (2018, May 16). *Determinants of genome editing outcomes: the impact of target and donor DNA structures*. Retrieved from https://hdl.handle.net/1887/62204

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/62204

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The following handle holds various files of this Leiden University dissertation:
http://hdl.handle.net/1887/62204

Author: Chen, X.
Title: Determinants of genome editing outcomes: the impact of target and donor DNA structures
Issue Date: 2018-05-16

Determinants of genome editing outcomes:

the impact of target and donor DNA structures

Xiaoyu Chen

ISBN: 978-94-6332-349-9

© Xiaoyu Chen, 2018, Leiden, the Netherlands. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without prior written permission of the author.

Cover: concept and design by Xiaoyu Chen. Digital rain, *Cas9* sequence.

The printing of this thesis was supported by the Virus and Stem Cell group, Department of Cell and Chemical Biology, Leiden University Medical Center.

Printed by GVO drukkers & vormgevers B.V. (www.proefschriften.nl)

Determinants of genome editing outcomes: the impact of target and donor DNA structures

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 woensdag 16 mei 2018
klokke 13:45 uur

door

Xiaoyu Chen
Geboren te Xingtai, Hebei, China
in 1990

Promotor: Prof. Dr. R.C. Hoeben
Co-promotor: Dr. M.A.F.V. Gonçalves

Leden promotiecommissie:
Prof. Dr. M. Tijsterman
Prof. Dr. A. Schambach (Hannover Medical School)
Prof. Dr. S.M. van der Maarel
Prof. Dr. R. Kühn (Max Delbrück Center for Molecular Medicine)
Dr. M.E. Ressing

The research presented in this thesis was performed at the Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands. This research was in part supported by the Dutch Prinses Beatrix Spierfonds (W.OR11-18), the French AFM-Téléthon (grant number 16621), and the China Scholarship Council.

谨以此书献给爸爸，妈妈和弟弟

以及恩师李春岩

To my mom, dad and brother,

and Chunyan Li

CONTENTS

General Introduction / **009**

Chapter 1 CRISPR-Cas9 Genome Editing, Five Years on / **013**

Chapter 2 In trans Paired Nicking Triggers Seamless Genome Editing without Double-stranded DNA Cutting / **031**

Chapter 3 Probing the Impact of Chromatin Conformation on Genome Editing Tools / **061**

Chapter 4 The chromatin Structure Differentially Impacts High-specificity CRISPR-Cas9 Nuclease Strategies / **081**

Chapter 5 The Chromatin Structure Governs Gene-editing Outcomes / **093**

Chapter 6 Engineered Viruses as Genome Editing Devices / **119**

Chapter 7 The Emerging Role of Viral Vectors as Vehicles for *DMD* Gene Editing / **143**

Summary and Concluding Remarks / **163**

Addendum Nederlandse Samenvatting / **170**
List of Publications / **173**
Curriculum Vitae / **174**
Acknowledgments / **175**

General introduction

The past few years have witnessed the explosive enthusiasm on genome editing. Scientists are better equipped with molecular tools to answer hard-to-answer questions, to cure incurable diseases, and to redefine pre-defined genetic code. Typically, genome editing is based on generating targeted chromosomal double-stranded DNA breaks (DSBs) in living cells. In response to such targeted DSBs, cells have evolved DNA repair mechanisms that can be exploited for genome editing purposes. The non-homologous end joining (NHEJ) pathway, which ligates broken chromosomal ends, can lead to gene knockouts after the incorporation of disruptive small insertions and deletions (indels); the homology-directed repair (HDR) pathway, in turn, copies the genetic information into the target sites from exogenous donor DNA templates that share homology with the target genomic sequence. Designer nucleases (a.k.a. programmable nucleases) are the molecular tools that generate the site-specific chromosomal DSBs necessary for genome editing. These tools include zinc-finger nucleases (ZFNs) and transcription activator-like effector (TALE) nucleases (TALENs), whose DNA recognition is mediated by protein-based zinc finger motifs and TALE repeats, respectively. Recently, designer nucleases based on prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 systems have been introduced. In contrast to their ZFN and TALEN predecessors, the CRISPR-Cas9-based platforms, consisting of RNA-guided nucleases (RGNs), do not require protein engineering efforts as their target DNA recognition is mediated by RNA-DNA hybridization, as opposed to protein-DNA interactions.

The adaptation of CRISPR-Cas9 nucleases to eukaryotic cells not only propels the development of genome editing, but also facilitates the engineering of a broad range of molecular tools for targeted transcriptional regulation, epigenetic remodeling and cellular imaging. Owing to their ease-of-use and cost-effectiveness, RNA-guided molecular tools based on CRISPR-Cas9 systems have, in a short time span, become the most widely used genome, transcriptome and epigenome manipulating platforms. Moreover, encouragingly, a number of clinical trials which make use of RGNs have recently been granted the green light by regulatory authorities. Examples include clinical trials aiming at the genetic engineering of chimeric antigen receptor T-cells and CD34+ cells from cancer patients and HIV-infected individuals, respectively. Nevertheless, one of the biggest concerns on the clinical translation of CRISPR-Cas9 nucleases is off-target activity and related potential adverse effects, e.g., deleterious or unpredictable mutations possibly compounded with oncogenic chromosomal translocations. Indeed, a number of studies have revealed that RGNs can cleave not only the chromosomal DNA at the intended target site, but also at off-target sequences bearing a varying number of mismatches. Therefore, whenever genome editing is meant to be applied in clinical settings, its specificity and accuracy

will be paramount and, as a result, potential deleterious effects caused by off-target activity should be thoroughly assessed. Although a great amount of efforts have been made to improve the safety profiles of RGNs, the fact that conventional genome editing strategies rely on DSB formation confers them a non-negligible risk. Of note, this risk is associated with unwanted mutations at not only, off-target, but also on-target sites.

In this thesis, **Chapter 1** reviews the fast-paced investigations aiming at improving the efficiency and specificity of genome editing procedures, covering the strategies based on protein engineering, gRNA design and donor DNA optimization. **Chapter 2** focuses on investigating a DSB-free genome editing principle that triggers seamless genetic modifications in human cells after the recruitment of the HDR pathway. This hereafter called in trans paired nicking strategy is based on inducing single-stranded DNA breaks (SSBs), or nicks, at genomic target sites and donor DNA templates. In this work, I present experimental evidence showing that, when compared to conventional DSB-dependent approaches, in trans paired nicking can improve the efficiency, specificity and accuracy of chromosomal DNA insertion at different human loci, including the dystrophin-encoding *DMD* gene, and the purported "safe harbor" loci *AAVS1* and *CCR5*. The extent of the site-specific genetic modifications varied from small restriction enzyme polymorphisms to whole transgene expression units. Taken together, these data *bode well* for the applicability of in trans paired nicking to settings in which non-disruptive and high-fidelity genetic manipulation of complex genomes is key.

Hitherto, it remains ill-defined how, and to what extent, the nuclear DNA architecture and the epigenetically regulated chromatin compaction affect genome editing endeavors based on designer nucleases. Hence, next to investigating DSB-free high-fidelity genome editing strategies, in this thesis, I have also probed the impact of alternate higher-order chromatin conformations on (i) the performance of TALENs and RGNs (**Chapter 3**); (ii) the performance of high-specificity designer nuclease constructions (**Chapter 4**); and (iii) the balance between NHEJ and HDR events during the gene editing interventions (**Chapter 5**). In this regard, I carried out genome editing experiments in complementary gain-of-function and loss-of-function cellular models in which the chromatin states of isogenic reporter loci are tightly regulated by small-molecule drug availability (i.e. doxycycline). The reporter genes in these cellular models are flanked by *TetO* sequences and are subjected to epigenetic regulation by the chimeric tTR-KRAB protein which, in the absence of doxycycline, binds to its cognate *TetO* recognition sites, inducing a local heterochromatic environment. This experimental setup has enabled the unbiased evaluation of the relative impact of "closed" heterochromatin and "open" euchromatin on both NHEJ- and HDR-mediated gene editing at isogenic target sequences. Gene editing experiments performed in these cellular models showed that NHEJ-mediated tar-

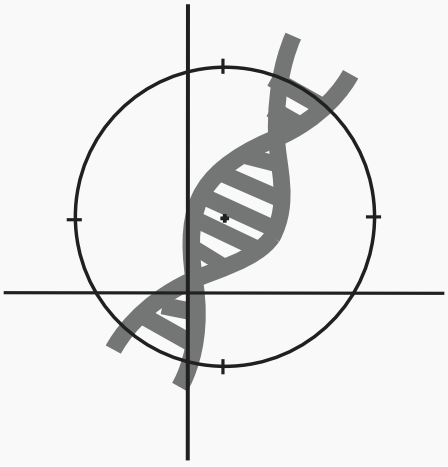
geted gene knockouts induced by TALENs and RGNs are significantly hampered at heterochromatic target sites (**Chapter 3**). **Chapter 4** further extends the investigations in **Chapter 3** by describing the differential impact of alternate higher-order chromatin conformations on different high-specificity RGN constructions, in particular those based on (i) offset RGN "nickase" pairs, (ii) improved Cas9 variants, (iii) truncated gRNAs and (iv) the ortholog *S. aureus* Cas9 protein. The experimental section of this thesis ends with **Chapter 5**, which investigates the impact of chromatin on genome editing outcomes resulting from the action of the HDR and NHEJ pathways with donor DNA templates of synthetic, non-viral and viral origins.

Another important aspect of genome editing is to ensure that the delivery of designer nucleases and donor DNA constructs is specific, robust and non-disruptive to the target cell's genome. This key "delivery" issue remains challenging especially when patient cells are to be modified either *ex vivo* or *in vivo*. Given their integration-defective character and well-established cell entry mechanisms, episomal viral vectors, i.e., engineered replication-defective viruses, might constitute ideal delivery vehicles for introducing in a targeted and efficient manner gene editing reagents into therapeutically relevant cells. In this regard, **Chapter 6** summarizes the principal characteristics, advantages and drawbacks of the main viral vector systems that are being adapted for genome editing purposes, namely, integration-defective lentiviral vectors (IDLVs), adeno-associated viral vectors (AAVs) and adenoviral vectors (AdVs). As a case-in-point, **Chapter 7** discusses the emerging role of viral vectors in repairing defective dystrophin-encoding genes in mouse models and human cells, whose mutations cause the lethal muscle-wasting X-linked disorder Duchenne muscular dystrophy (DMD).

In conclusion, by investigating the interaction between different types of nucleases (i.e. nicking versus cleaving), donor DNA structures and target chromatin environments, this thesis provides important insights into how to improve the three crucial parameters of genome editing: efficiency, specificity and fidelity. Hence the work presented in this thesis expand the range of possibilities for high-fidelity genetic manipulation of human cells.

Chapter 1

CRISPR-Cas9 Genome Editing, Five Years on



Up until about five years ago, when speaking of genome editing using designer nucleases, scientists were mostly referring to transcription activator-like effector nucleases (TALENs) and zinc-finger nucleases (ZFNs); now RNA-guided nucleases (RGNs) based on clustered regularly interspaced short palindromic repeats (CRISPR) systems are, undoubtedly dominating the field [1]. Native type II CRISPR-Cas9 nucleases from *S. pyogenes* were the first to be adapted into genome editing tools [2-4] and, similarly to other prokaryotic CRISPR systems, confer adaptive immunity by detecting, recording and degrading foreign DNA from invading viruses and plasmids [5, 6]. Soon after the discovery in 2012 that *S. pyogenes* CRISPR-Cas9 nucleases could cleave DNA in an RNA-guided manner [5, 6], researchers hijacked this defense system and turned it into a powerful tool to edit the genomes of eukaryotic cells, in a targeted and permanent basis [2-4]. These genome engineering technologies have wide implications for basic and applied research and bioethics, especially in the context of genetically modifying human somatic and germ cells. With the introduction and ongoing refinement of RGNs, it seems that the scientific community is on track to tackle the root cause of genetic and infectious diseases. Nevertheless, RGNs suffer from, amongst other shortcomings, off-target activities and relatively low homology-directed repair (HDR) efficiency. To overcome these hurdles, and hence unleash the full potential of RGN-based therapies, researchers are rationally engineering Cas9 variants, *de novo*-isolating Cas9 proteins, fine-tuning gRNA and donor DNA structures, and devising improved protocols for obtaining specific genome editing outcomes. This chapter reviews these efforts that, when combined, might overcome the main limitations of current genome editing technologies.

Genome editing in a nutshell

Efficient genome editing relies on designer nucleases, which, by making site-specific chromosomal double-stranded DNA breaks (DSBs), trigger cellular DNA repair pathways to establish specific genome editing outcomes., e.g., targeted gene knockouts and knock-ins [1, 7, 8]. The non-homologous end joining (NHEJ) can knockout *cis*- or *trans*-acting elements after the introduction of small insertions and deletions (indels); the homology-directed repair (HDR) can knock-in new genetic information using exogenous DNA as surrogate DSB-repairing templates (donor DNA) [1, 7]. It is noteworthy mentioning that, in mammalian cells, DNA repair often favors NHEJ over HDR, with the HDR machinery largely restricted to the G2/S phases of the cell cycle [9, 10]. Among the designer nucleases used in genome editing, ZFNs and TALENs share a similar architecture in that a sequence-specific DNA-binding protein domain is fused to the nuclease domain of the type IIS FokI restriction enzyme. Moreover, ZFNs and TALENs work in pairs, cleaving their target DNA after the *in situ* dimerization and ensuing catalytic activation of their FokI domains [1, 7]. The vertebrate Cys2-His2 zinc-finger motif is the basic unit of the DNA-binding domain of ZFNs

with each zinc-finger typically recognizing a specific DNA triplet. Thus, assembling 4-6 zinc-finger motifs per ZFN monomer generally provides enough target site specificity within complex eukaryotic genomes [11-16]. TALENs recognize DNA through an array of repeated domains of TALE proteins found in certain phytopathogenic bacteria, e.g., *Xanthomonas sp.* [17, 18]. Each TALE repeat is composed of 33-35 conserved amino acids with variable residues dubbed repeat variable di-residues at positions 12 and 13, recognizing a specific nucleotide on the DNA double-helix [17, 18]. Researchers have been exploiting ZFNs and TALENs not only for generating transgenic animal models [19], but also for treating diseases, including AIDS [20], hemophilia B [21] and certain leukemia [22]. Nevertheless, many laboratories flinch from using ZFNs and TALENs because making these protein-guided nucleases requires more expertise and higher costs than those of RGNs.

RGNs from the prototypic *S. pyogenes* type II CRISPR-Cas9 system and, more recently, those from other species, are offering unprecedented speed and versatility with which robust genome editing tools can be built [1, 7]. The RGNs from *S. pyogenes* consist of a fixed Cas9 nuclease (SpCas9), and two RNAs, i.e., a sequence-specific CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). Together these two RNAs form a crRNA:tracrRNA duplex that binds the SpCas9 nuclease and directs it to a target site consisting of an NGG triplet called protospacer-adjacent motif (PAM) and a 20-nt long sequence complementary to the 5′ end of the crRNA (spacer) [5, 6, 23]. The recognition of the NGG triplet by the PAM-interacting domain of SpCas9 is followed by double helix melting, local RNA-DNA hybridization, R-loop elongation and PAM-dependent allosteric activation of the HNH and RuvC-like nuclease domains [24]. This series of events lead to the generation of a blunt DSB mostly positioned 3-bp distal from the PAM [5, 6, 24]. The initial adaptation of the native *S. pyogenes* CRISPR-Cas9 nuclease system into a powerful genome editing tool involved expressing human codon-optimized *Cas9* ORF and single guide RNAs (sgRNAs) formed by the fusion of crRNA and tracrRNA moieties that mimic the original crRNA:tracrRNA structure [2, 4]. Hence, by simply swapping the spacer sequence in the sgRNA, researchers can target virtually any genomic sequence with unprecedented flexibility. Depending on the target cells —cultured *in vitro* or located in a living organism—RGNs can be delivered as ribonucleoproteins (RNPs) or encoded in mRNA, plasmid DNA or viral vector genomes; and depending on the purposes, they can be designed to induce (i) disruption of specific sequences (knockout), (ii) gene knock-ins, (iii) targeted DNA deletions, (iv) directed chromosomal translocations and (v) single nucleotide substitutions.

Notable bottlenecks of CRISPR-Cas9 technologies

Admittedly, there are no perfect designer nucleases. As for ZFNs and TALENs, RGNs can cleave at unintended off-target sites, which can only be partially identified

through *in silico*-guided and genome-wide assays [1, 25]. Given the complexity and very large size of the human genome (~3 billion base pairs per haploid genome), the generation of off-target DSBs poses a non-negligible risk, which might knockout essential genes or mutate cancer-associated genes [26-30]. These outcomes must especially be avoided when the goal is to develop genetic therapies that exploit the NHEJ or the HDR pathways. In the context of gene knockout approaches, the unpredictable indel footprints from NHEJ might produce aberrant products, e.g., truncated or misfolded proteins or proteins displaying immunogenic epitopes. In the context of genome editing strategies involving multiple DSBs at different loci, chromosomal rearrangements might arise contributing to genomic instability [31]. Regarding HDR-based genome editing, RGNs, next to faithfully integrating the exogenous DNA at one allele, often also disrupt the other allele due to the prevalence of NHEJ over HDR in mammalian cells [32]. Moreover, HDR-mediated chromosomal DNA insertion triggered by site-specific DSB formation often is not efficient enough (< 1%), especially when donor templates encompass entire transcription units stretching several kilobases. Finally, the native chromatin environment in eukaryotic cells poses yet another hurdle compromising the efficiency and predictability of genome editing efforts [33, 34].

To address the aforementioned issues, many researchers are investigating a broad range of strategies that are as diverse as engineering or mining for new nucleases, redesigning sgRNA moieties and testing new donor DNA structures aiming at achieving efficient and accurate genetic manipulations in complex (epi)genomes.

Isolating and engineering new Cas9 proteins

Although the SpCas9 PAM sequence NGG occurs in the human genome every 8 bp on average, it can nonetheless constitute a limitation, especially when the genomic sequences of interest have a high A/T content. Using a selection-based SpCas9 system in bacteria, Kleinstiver et al. isolated the VQR and VQER variants of SpCas9, each of which recognizing the PAMs NGAN and NGCG, respectively [35]. Importantly, their directed evolution system has also yielded the D1135E SpCas9 variant which displays a superior discriminating profile between canonical NGG and cryptic NGA PAMs than that of the wild-type SpCas9 protein [35].

Multiple CRISPR-Cas systems isolated from different species have been adapted for genome editing in human cells, including those from *S. aureus* and *Acidaminococcus sp.*. Often, the resulting RGNs display an improved or complementary set of attributes when compared to the SpCas9. For instance, in contrast to the 4.1-kb ORF of the SpCas9 nuclease (1368 amino acids), the ORF coding for the *S. aureus* Cas9 (SaCas9) protein (1053 amino acids) spans only ~3.2 kb, making it easily packaged within the limited confinements of commonly-used viral vectors, e.g., adeno-associated viral vectors [36]. In addition, Cas9 orthologs offer alternative specificity profiles and

genomic targeting ranges owing to their different PAM sequences, e.g., SaCas9 has as PAM the NNGRRT motif instead of NGG [36]. To further increase the versatility of SaCas9-based RGNs, Kleinstiver et al. generated the KKH SaCas9 variant, which recognizes the degenerate PAM sequence NNNRRT and, as a result, has broadened the genomic coverage of these RGNs [37]. Different from the RGNs based on type II CRISPR-Cas9 systems of *S. pyogenes* and *S. aureus*, those adapted from type V CRISPR-Cpf1 systems recognize T-rich PAM sequences, generate staggered DSBs with 5′-overhangs, display enhanced specificities and do not have a tracrRNA component [38]. Similarly to the development of Cas9 proteins with altered PAMs, using structure-guided mutagenesis, Gao and colleagues engineered mutants of *Acidaminococcus sp*. Cpf1 (AsCpf1) which recognize the PAMs TYCV and TATV instead of the native PAM sequence TTTV [39].

In parallel to modifying the target range of RGNs, recent research is also leading to improvements in reducing RGN off-target activities. Previous studies on the development of ZFNs and TALENs have set the stage for optimizing the specificity of RGNs by mimicking the *modus operandi* of ZFNs and TALENs, which encompass FokI nuclease domain dimerization [13, 40]. Tsai et al. and Guilinger et al. fused the FokI nuclease domain to the N-terminus of a catalytically inactive SpCas9 protein to form FokI-dCas9 monomers (where "d" stands for "dead") [41, 42]. The binding of a pair of these monomers to their bipartite target sequences forms a dimeric RNA-guided FokI nuclease (RFN), offering a higher specificity profile when compared to conventional RGNs as the target sequences of RFN span up to 44 bp. Despite of its improved specificity, the RFN approach has limited genomic DNA coverage due to their strict gRNA design and spacing requirements. In particular, RFNs require a so-called PAM-out gRNA design with their activities peaking within a 14-17 bp spacing between the two hemi-nucleases [41, 42]. These limitations, together with the fact that RFNs are larger than conventional RGNs, are keeping this approach at the proof-of-principle stage. The nuclease domains of SpCas9, HNH and RuvC, cut the target and non-target strands, respectively. Mutating catalytic residues in each of these nuclease domains yields the RuvC mutant D10A and the HNH mutant H840A "nickases", which cleave the target and non-target strands, respectively [6]. Delivering into cells a Cas9 "nickase" together with two different gRNAs targeting opposite DNA strands of a bipartite recognition sequence induces a DSB owing to the local coordinated action of both RGN "nickases". Should either of the nicking complexes cleave at off-target sites, the resulting SSBs are, for the most part, faithfully repaired. Thus, this confers an overall high specificity to the dual nicking approach. When compared to conventional RGNs, this dual nicking strategy can reduce off-target activities (up to 1,500-fold) [43, 44]. Similar to RFNs, the relative positions and spacing of the two gRNAs are important for efficient target DNA cleavage. Usually, dual RGN "nickases" yielding 5′-overhangs and a spacing between the 5′ ends of the two gRNAs

between -4 and 20 nt result in the highest DSB formation activities [43, 44]. Of note, SaCas9 "nickases" bearing a D10A or an N580A mutation were also compatible with the dual nicking strategy, although in this case, the optimal distance between the gRNAs seems to lie between 0 and 125 bp [45]. One drawback of the dual nicking strategy is that, often, there is a trade-off in which the gains in specificity are accompanied by a decrease in efficiency [43-45].

To obviate the need for a pair of suitable gRNAs for dual nicking to increase the specificity of RGNs, researchers have generated high-specificity SpCas9 variants through structure-guided protein engineering involving mutations-to-alanine of specific DNA-interacting residues [46, 47]. They reasoned that the resulting reduction in the RGN-DNA binding energies would be adequate to cleave the target site but insufficient to cleave at off-target sites [46, 47]. Slaymaker et al. reported enhanced specificity of the SpCas9 variants eSpCas9(1.0) and eSpCas9(1.1) containing triple mutations (i.e., K810A/K1003A/R1060A and K848A/K1003A/R1060A, respectively) at the positively charged nt-groove of SpCas9, which plays a role in stabilizing the non-target DNA strand [47]. Kleinstiver et al. generated in turn quadruple mutant SpCas9-HF1 (i.e., N497A/R661A/Q695A/Q926A). These mutations-to-alanine are thought to attenuate hydrogen bonds between SpCas9 and the phosphate backbone of the target DNA strand [46]. Both strategies significantly reduce off-target chromosomal DNA cleavage while retaining on-target activities similar to those of wild-type SpCas9 for the majority of target sequences tested [46, 47]. However, it is noteworthy mentioning that the eSpCas9(1.1) and SpCas9-HF1 variants are less efficient than wild-type SpCas9, when the gRNAs contain a 5'-mismatched nucleotide (e.g. 21-mer spacer with a 5' end "G" extension) or are truncated (i.e. < 19- to 20-mer spacers) [33, 48]. Interestingly, the SpCas9 variant HeFm2SpCas9, harboring combinatorial mutations from eCas9(1.1) and SpCas9-HF1, albeit more specific than their parental proteins, suffers from low activity at most target sites tested [49]. More recently, Chen et al. showed that the specificity gains of eSpCas9(1.1) and SpCas9-HF1 were likely due to a failure in the conformational change of the REC3 domain necessary for HNH catalytic activation at off-target sites [50]. Based on this information, they engineered a SpCas9 protein containing the REC3 mutations N692A/M694A/Q695A/H698A, and showed that this so-called HypaCas9 has comparable or higher specificity than SpCas9-HF1 and eSpCas9(1.1) [50].

To circumvent DSB formation and the undesirable NHEJ-derived mutations, researchers are developing strategies that resort to the induction of SSBs instead of DSBs to facilitate targeted genome editing [32, 51, 52]. Komor et al. and Nishida et al. introduced the base editing concept, which ultimately replaces one base pair for another [51, 52]. The rationale behind base editors is that the cytidine in a C:G pair can be converted to a U:G pair by cytidine deaminases, e.g., APOBEC1 [51] and AID [52]. After mismatch repair the U:G becomes a U:A pair, which finally turns into a

T:A pair after DNA replication. The first generation of base editors made by Komor et al. has the cytidine deaminase APOBEC1 fused to the dSpCas9 (BE1). To avoid removal of uracil by uracil glycosylases, a uracil glycosylase inhibitor (UGI) was added to their second generation base editor (BE2). To increase the base editing efficiency, the dSpCas9 was replaced by a SpCas9 "nickase", yielding a third generation base editor (BE3) consisting of a fusion between APOBEC1, a UGI and the SpCas9 D10A "nickase" scaffold. BE3 generates an SSB in the G-containing strand converting any C:G base pair to a T:A (C to T) within a 5-nt window about 12 bp away from the PAM [51]. More recently, a BE3 mutant named YEE-BE3 with a weaker cytidine deaminase activity has narrowed the peak of C-to-T transitions from a 5-nt to a 2-nt window [53]. In addition, base editors containing as scaffolds SpCas9 variants or SaCas9 have either expanded the PAM requirements of base editors [54] or improved their specificity [55]. With a fine-tuned design, the fourth generation base editor, BE4, has larger linkers between the SpCas9 "nickase", APOBEC1 and UGI, and incorporates a second copy of UGI [54]. This construction enriches for the intended C-to-T transitions and reduces unwanted conversions likely mediated by uracil glycosylases. Other BE3 and BE4 base editors (e.g. BE3-Gam, SaBE3-Gam, BE4-Gam, and SaBE4-Gam) incorporate the bacteriophage Mu Gam protein which purportedly, by binding to DSBs, reduces the residual levels of nick-derived chromosomal breaks [54]. Finally, The recent development of adenine base editors (ABEs) nicely complements their BE counterparts as these new based editors can convert A:T to G:C pairs (A to G) in a highly efficient manner [56, 57]. Theoretically, with current base editors, many pathogenic single base mutations can be corrected and, in addition, ATG start codons can be disabled for targeted and precise gene knockouts.

Redesigning and modifying gRNAs

sgRNAs have secondary structures forming hairpins and stem loops which mimic the architecture of native gRNAs consisting of the sequence-specific crRNA (42 nt) and the scaffolding tracrRNA (89 nt) [2, 6]. The 5'-terminal 20 nts of the sgRNA (spacer) and the Cas9-interacting scaffold function as essential elements in sgRNA molecules [2]. Truncated gRNAs (tru-gRNAs) displaying 17-19 mer spacers instead of the typical 20 mer spacers, have a weaker DNA-binding potential which, presumably, biases the cleaving activity to fully complementary target sites in detriment of mismatched (off-target) sequences. Indeed, Fu et al. showed that RGNs bearing tru-gRNAs and wild-type SpCas9 can display greatly decreased off-target effects (up to 5,000-fold) while maintaining their on-target activity [58].

The most commonly used sgRNA structure has (i) a shorter crRNA:tracrRNA duplex region next to the spacer sequence when compared to that of the native crRNA:tracrRNA duplex, potentially impairing stable Cas9:gRNA complex assembly, and (ii) a continuous stretch of four uridines, which might serve as a premature

RNA polymerase III termination signal. On the basis of this information, Chen et al. and Dang et al. optimized the gRNA scaffold by extending the duplex region next to the 5′ spacer region by 5 bps, and showed significantly enhanced dSpCas9-based chromosomal imaging and SpCas9-mediated gene knockout, respectively [59, 60]. By combining the sgRNA duplex extension with the disruption of the uridine stretch, Dang and coworkers demonstrated higher *CCR5* and *CD4* gene knockout frequencies when compared to those achieved by the original sgRNA scaffold [60]. Considering that these optimized scaffolds seem to yield more stable and full-length sgRNAs, they might also be compatible with different (epi)genome editing tools, e.g., Cas9 variants, dCas9-based transcriptional regulators and base editors.

Adding functional RNA elements into sgRNA secondary structures can repurpose RGN-based tools, without diminishing their DNA targeting proficiency [43, 61]. For instance, hairpin aptamers can be inserted into the sgRNA scaffold at the tetraloop and lower stems, so that sgRNAs bind the MS2 bacteriophage coat protein. When such aptamer-containing sgRNAs are delivered together with Cas9 and MS2 fused to effector domains (e.g. VP64 transcription activator or Tet1-CD demethylation domains), potent transcription activation and epigenetic remodeling of target loci can be achieved [62, 63].

Currently, RNA polymerase III (Pol III) promoters are the most widely used elements for sgRNA synthesis as they evolved for expressing short, unprocessed, transcripts in eukaryotic cells. Although highly active, Pol III promoters require a G (e.g., U6 and 7SK promoters) or an A or a G (e.g., H1 promoter) to efficiently initiate transcription, therefore the most suited sgRNAs have spacers starting with a G or an A at their 5′ terminus [64, 65]. Of note, there is an impairment on the activity of the eSpCas9(1.1) and SpCas9-HF1 variants when the 5′ terminal G does not hybridize to the target site [33, 48]. Although RNA polymerase II (Pol II)-driven transcripts are subjected to complex downstream processes and nucleus-to-cytoplasm translocation, they offer the possibility for conditional or tissue-specific expression or for the expression of multiple sgRNAs from a single template. In this regard, Nissim and coworkers designed Pol II expression units containing the sgRNA coding sequence flanked by recognition sites for the RNA endonuclease Csy4 (Csy4-gRNA-Csy4), so that single or multiple sgRNAs can be released after Csy4 cutting [66]. The Csy4-gRNA-Csy4 unit was also inserted into an artificial intron construct with splicing sites and branching point. Hence, after splicing, the Csy4-gRNA-Csy4 module can be set free as an intron [66]. However, inserting the sgRNA sequence alone (i.e., without the flanking Csy4 sites and Csy4 expression) into the artificial intron construct did not yield stable sgRNA due to rapid degradation [66]. To avoid the concomitant delivery of the RNA endonuclease Csy4, the sgRNA sequence can be put after a 5′ hammerhead ribozyme (HH) or in-between a 5′ HH and a 3′ HDV ribozymes (HH-gRNA-HDV), which generate mature sgRNAs through self-cleaving [66-68]. Of note,

however, Yoshioka et al. showed that ribozyme-flanked sgRNAs expressed from the RNA Pol II CAG promoter (CAG-HH-gRNA-HDV) induced less genome editing events than those from the commonly used U6-driven sgRNA [68]. Alternatively, Xie et al. hijacked the endogenous tRNA maturation mechanism, and constructed a Pol III promoter-driven tRNA-sgRNA tandem array, which, after cleavage by endogenous RNAses, yielded functional sgRNAs in rice [69]. Follow-up studies confirmed that sgRNAs produced from the tRNA-gRNA module could achieve efficient genome editing in *Y. Lipolytica*, maize and, more recently, in *Drosophila* [70-72].

Pre-assembling Cas9 and gRNA ribonucleoprotein complexes *in vitro* before their delivery into target cells bypasses chromosomal integration of foreign DNA encoding these RGN components, and ensures a more transient activity window of the nucleases [73, 74]. The gRNAs are usually made from *in vitro* transcription or by solid-phase chemical synthesis [73, 74]. Unlike *in vitro* transcribed gRNAs, chemically synthesized gRNAs bypass the difficult-to-upscale enzymatic reactions. Lately, several groups extensively investigated the chemical modification of the gRNA at different positions of the nucleotide (base, ribose sugar and phosphate group), to increase their stability and genome editing efficiency [75-78]. Hendal et al. showed that sgRNAs modified with 2′-O-methyl, 2′-O-methyl-3′-phosphorothioate or with 2′-O-methyl-3′-thioPACE at both ends of the sgRNA induced robust genome editing at the *CCR5* and *HBB* loci in the CD34+ hematopoietic stem and progenitor cells [75]. In addition, Rahdar and co-workers systematically screened crRNAs modified at different positions with chemical substitutions in their phosphate and/or sugar groups. They identified a shortened, yet equivalently functional, 29 nt-long crRNA containing 2′-fluoro (2′-F) and 2′-S-constrained ethyl (2′-cEt) at their 5′ and 3′ ends, respectively [76]. Interestingly, in addition to improving the stability and reducing the size of gRNA components, chemical modifications are also paying off in terms of generating reagents with enhanced specificity. A recent study demonstrated that 2′-O-methyl-3′-phosphonoacetate modifications at positions 5 or 11 of sgRNAs can remarkably reduce off-target activity while maintaining on-target cleavage [77]. Instead of substituting chemical groups in the ribose-phosphate backbone, Lee and co-workers added large groups at the termini of crRNAs (e.g. 5′-Rhodamine and 5′-Amine) to endow them with new functions, i.e., fluorescence for enriching edited cells [78].

Manipulating donor DNA templates

As aforementioned, when partnered with designer nucleases, donor DNA molecules can act as DSB-repairing templates to precisely incorporate customized DNA changes into specific genomic positions through the HDR pathway [1]. Regardless of their foreign nucleic acid composition, conventional donor DNA templates have sequences identical to those framing a designer nuclease target site (“homology

arms"). These "homology arms" serve as strand invasion regions that allow for homology-directed insertion of foreign DNA, whose size can vary from single base pairs to whole transgenes [1]. For the targeted insertion of large transgenes, the donor DNA is often provided in recombinant plasmids or viral vector genomes. Yet, regardless of the delivery method, HDR occurs mainly during the late G2 and S phases of the cell cycle, and is particularly inefficient in human cells, such as hematopoietic stem/progenitor cells and iPSCs [1, 9, 10]. Moreover, the competing and constantly available NHEJ pathway leads to a large fraction of heterozygous edited alleles due to the concurrent action of HDR and NHEJ, or NHEJ alone.

Studies on the design and testing of different types of donor DNA molecules unveiled attractive research avenues to increase the efficiency, specificity and/or accuracy of HDR-based genome editing [1]. Holkers and coworkers demonstrated that the donor templates delivered in protein-capped adenoviral vector genomes greatly enhance the specificity and fidelity of HDR-based genome editing [79]. By using a "double-cut" donor DNA design in which the "homology arms" are flanked by sgRNA target sites, Zhang et al. achieved 8 % and 12 % transgene knock-ins in human iPSCs at the *CTNNB1* locus with 300 bp and 600 bp "homology arms", respectively [80]. These targeted DNA insertion frequencies were 2- to 5-fold higher than those achieved by the standard covalently-closed plasmid templates [80]. Chen et al. repoted a similar effect after comparing "double-cut" with standard donor plasmids at the *DMD* and *AAVS1* loci in human cells [32]. To bypass the HDR pathway, researchers are also looking into "double-cut" donors containing no or short "homology arms" (e.g., 10-40 bps) for DNA insertions, which are achieved by NHEJ and microhomology-mediated end joining (MMEJ) pathways instead [81-83]. By using RGNs together with homology-free "double-cut" donors, He et al. showed up to 20 % and 1.7 % of a 4.6-kb transgene insertion at the *GAPDH* locus in LO2 cells and ESCs, respectively [83]. Suzuki et al. also applied the homology-free "double cut" donors and inserted different expression units in post-mitotic mouse neurons, both *in vitro* and *in vivo* [84]. Alternatively, "double-cut" donors with short "homology arms" were efficient substrates for MMEJ-mediated transgene insertion in zebrafish [85] and, more recently, in mice [82]. However, it is of note that free DNA termini generated from "double-cut" donors increase the risk of random and/or foreign DNA insertions resulting from imprecise recombination processes [32, 79, 80].

Albeit at low levels, SSB-induced HDR using designer "nickases" can yield targeted and accurate chromosomal insertion of foreign DNA in mammalian cells without the attendant catalytic induction of DSBs [32, 86, 87]. Recently, Chen and co-workers have developed an efficient SSB-induced HDR approach, which can place large DNA segments into specific genomic target sites without provoking the mutagenic NHEJ [32]. The key aspect to this approach, named in trans paired nicking, consists

of combining a Cas9 "nickase" with a modified donor DNA in which the "homology arms" are framed by the sgRNA recognition site [32]. This arrangement assures concomitant SSB formation at target and donor sequences, generating homologous recombination DNA substrates as postulated by Holliday in 1964 [88]. The in trans paired nicking not only increased HDR at the human *AAVS1*, *CCR5* and *DMD* loci in different cell types, including iPSCs, but also facilitated multiplexing HDR [32].

Single-stranded oligodeoxyribonucleotides (ssODNs) are commonly used donors for small genomic modifications, e.g., point mutations, polymorphisms or short protein-tagging motifs. Typical ssODN donors are symmetrical in that they harbor similarly sized "homology arms" at both ends. Recent studies have shown that, when compared to symmetric ssODNs, asymmetric ssODNs can induce more efficient RGN-assisted genome editing [89]. Specifically, optimal ssODN donors should be complementary to the sgRNA non-target strand and encompass shorter homology at the PAM distal end [89]. Biophysical data indicate that this configuration favors donor-target hybridization owing to the generation of a 3′-ended flap on the non-target strand after DNA cutting [89]. Liang et al. confirmed the superiority of asymmetric ssODNs and, in addition, tested short double-stranded ODNs (dsODNs) consisting of annealed ssODNs displaying either 3′ or 5′ 30-nt overhangs [90]. They showed that dsODN donors with 3′ overhangs lead to higher frequencies of RGN-induced genome editing than those achieved with asymmetric ssODNs or with dsODNs containing 5′ overhangs [90].

Similarly to sgRNA components, donor DNA templates are also substrates for chemical modifications to improve the performance of genome editing. For instance, chemically modifying the terminal nucleotides of ssODNs with phosphorothioate increased the efficiency of genome editing by 2-3 fold [91]. In another study, Mang et al. introduced the CAB system, short for Cas9-Avidin-Biotin ssDNA, in which Avidin is fused to Cas9 and binds Biotin-conjugated ssODNs inside cells [92]. Equally relying on high-affinity Biotin-Avidin interactions, Carlson-Stevermer et al. developed the S1mplex system, where Biotin-conjugated ssODNs are tethered *in vitro* to RGNs containing sgRNAs displaying a Streptavidin-binding S1m aptamer [93]. Both the CAB and S1mplex systems increased the frequencies of HDR and the ratio of HDR to NHEJ in mammalian cells [92, 93].

Tweaking experimental conditions for improving genome editing outcomes

Besides engineering novel genome editing components, e.g., the above-described Cas9 proteins, sgRNA structures and donor DNA templates, researchers are also optimizing experimental protocols to improve the efficiency and predictability of genome editing procedures. A simple methodology involves modulating the temper-

ature at which the designer nucleases operate. When compared to the normal 37°C mammalian cell culture condition, ZFN and TALEN induced higher gene knockout frequencies under a transient, 30°C hypothermic period [94, 95]. In contrast, at least in certain mammalian cell lines, transfection of RGN-encoding plasmids at 39°C triggered significantly higher targeted gene knockout frequencies, presumably owing to increased sgRNA expression in these lines [96]. Other protocols entail enriching for cells that have undergone genome editing - by co-delivering reporter units that reflect the expression of designer nucleases [97], or by co-targeting an endogenous gene (i.e., *ATP1A1*) whose loss-of-function is selectable [98].

Steering DNA repair machineries in favor of HDR over NHEJ is also being exploited when targeted gene knock-ins are the desired genome editing outcomes. One can distinguish two main strategies: genetic manipulations or pharmacological interventions. Regarding the former, Chu and colleagues expressed short hairpin RNAs to silence the genes encoding the NHEJ factors KU70, KU80 or DNA ligase IV or co-expressed adenovirus serotype-4 E1B55K and E4orf6 proteins to induce proteasomal degradation of DNA ligase IV [99]. These and other authors have also applied the small-molecule DNA ligase IV inhibitor Scr7 to bias the repair of site-specific DSBs towards the HDR pathway in mammalian cell lines [99, 100], as well as in fertilized mouse and rat zygotes [100, 101]. Robert and co-workers have equally applied a pharmacological approach to dampen the rate of NHEJ and increase the frequency of HDR, but used the DNA-PKcs inhibitors NU7441 and KU-0060648 instead [102]. Next to testing NHEJ inhibitors, researchers are also evaluating HDR-enhancing agents as well. For instance, RS-1, a molecule previously shown to activate RAD51, which play a role in homology search and strand exchange, increased knock-ins by 2- to 5-fold in cell lines and pronuclear stage rabbit embryos [103]. This enhancement on HDR was observed after DSB formation by RGNs and TALENs [104, 105]. By screening a library of roughly 4,000 small molecules, Yu and coworkers found that the β3-adrenergic receptor agonist L755507 stimulates HDR-mediated *Nanog* targeting by 3-fold in murine PSCs [106]. Alternatively, timed delivery of *in vitro* assembled ribonucleoprotein RGNs together with ssODN donors to cells that are synchronized at the cell cycle G1/S phase border by aphidicolin or at the G2/M phase by nocodazole, enhanced HDR-mediated genome editing up to 6-fold when compared with unsynchronized cells [107]. Of note, whether the use of specific genetic and/or pharmacological agents for modulating HDR/NHEJ ratios have long-term pleiotropic effects on target cells or organisms will require further investigations.

Concluding remarks

The ongoing discovery of Cas9 orthologs combined with structure-guided and directed evolution-based protein screens are yielding an ever-increasing number of Cas9 variants with new properties, e.g., strand-specific cleaving activities, alternative

PAM requirements and higher target DNA specificities. In addition, albeit not suited for large genomic DNA changes, base editors promise precise, DSB-free, genome editing. Next to isolating and engineering new Cas9 proteins, reshaping the native gRNA architecture and chemically modifying the gRNA composition are improving the activity and specificity of RGNs. By the same token, manipulating the structure and composition of donor DNA molecules is paying off in terms of enhancing the efficiency and specificity of HDR-based genome editing. To expand and complement hypothesis-driven studies for identifying HDR-enhancing agents, high-throughput screenings of small-molecule drug libraries should yield new compounds for modulating DNA repair pathways. Finally, sensitive and unbiased genome-wide assays will become ever-more important to thoroughly access the global mutation levels upon the application of the constantly improved genome editing tools.

References

1. Maggio, I. and Goncalves, M.A. (2015) Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 33 (5), 280-91.
2. Cong, L. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339 (6121), 819-23.
3. Jinek, M. et al. (2013) RNA-programmed genome editing in human cells. Elife 2, e00471.
4. Mali, P. et al. (2013) RNA-Guided Human Genome Engineering via Cas9. Science 339 (6121), 823-826.
5. Gasiunas, G. et al. (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 109 (39), E2579-E2586.
6. Jinek, M. et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337 (6096), 816-21.
7. Gaj, T. et al. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31 (7), 397-405.
8. Rouet, P. et al. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A. 91 (13), 6064-8.
9. Krejci, L. et al. (2012) Homologous recombination and its regulation. Nucleic Acids Res. 40 (13), 5795-818.
10. San Filippo, J. et al. (2008) Mechanism of eukaryotic homologous recombination. Annu Rev Biochem. 77, 229-57.
11. Bitinaite, J. et al. (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A. 95 (18), 10570-5.
12. Fairall, L. et al. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature 366 (6454), 483-7.
13. Kim, Y.G. et al. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 93 (3), 1156-60.
14. Pabo, C.O. et al. (2001) Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem. 70, 313-40.
15. Segal, D.J. and Barbas, C.F., 3rd (2001) Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. Curr Opin Biotechnol. 12 (6), 632-7.
16. Wolfe, S.A. et al. (2000) DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct. 29, 183-212.
17. Boch, J. et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326 (5959), 1509-12.
18. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by

TAL effectors. Science 326 (5959), 1501.
19. Tesson, L. et al. (2011) Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol. 29 (8), 695-6.
20. Tebas, P. et al. (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N Engl J Med. 370 (10), 901-10.
21. Sharma, R. et al. (2015) In vivo genome editing of the albumin locus as a platform for protein replacement therapy. Blood 126 (15), 1777-84.
22. Qasim, W. et al. (2017) Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. Sci Transl Med. 9 (374), eaaj2013.
23. Deltcheva, E. et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.
24. Nishimasu, H. et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156 (5), 935-49.
25. Hendel, A. et al. (2015) Quantifying On and Off-Target Genome Editing. Trends in Biotechnology 33 (2), 132-140.
26. Cho, S.W. et al. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24 (1), 132-41.
27. Cradick, T.J. et al. (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. 41 (20), 9584-92.
28. Fu, Y. et al. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 31 (9), 822-6.
29. Kuscu, C. et al. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol. 32 (7), 677-83.
30. Lin, Y. et al. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. 42 (11), 7473-7485.
31. Ousterout, D.G. et al. (2015) Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nat Commun. 6, 6244.
32. Chen, X. et al. (2017) In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. Nat Commun. 8, 657.
33. Chen, X. et al. (2017) The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. Mol Ther Nucleic Acids. 8, 558-563.
34. Chen, X. et al. (2016) Probing the impact of chromatin conformation on genome editing tools. Nucleic Acids Res. 44 (13), 6482-92.
35. Kleinstiver, B.P. et al. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523 (7561), 481-5.
36. Ran, F.A. et al. (2015) In vivo genome editing using Staphylococcus aureus Cas9. Nature 520 (7546), 186-91.
37. Kleinstiver, B.P. et al. (2015) Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 33 (12), 1293-1298.
38. Zetsche, B. et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163 (3), 759-71.
39. Gao, L. et al. (2017) Engineered Cpf1 variants with altered PAM specificities. Nat Biotechnol. 35 (8), 789-792.
40. Christian, M. et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186 (2), 757-61.
41. Guilinger, J.P. et al. (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat Biotechnol. 32 (6), 577-582.
42. Tsai, S.Q. et al. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. 32 (6), 569-76.

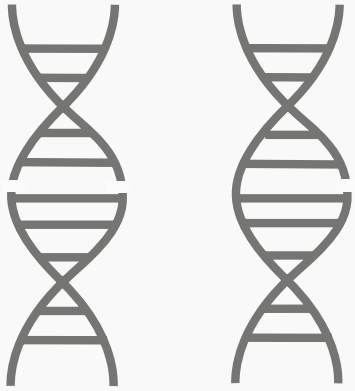
43. Mali, P. et al. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol. 31 (9), 833-8.
44. Ran, F.A. et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154 (6), 1380-9.
45. Friedland, A.E. et al. (2015) Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome Biol. 16, 257.
46. Kleinstiver, B.P. et al. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529 (7587), 490-5.
47. Slaymaker, I.M. et al. (2016) Rationally engineered Cas9 nucleases with improved specificity. Science 351 (6268), 84-8.
48. Kim, S. et al. (2017) Rescue of high-specificity Cas9 variants using sgRNAs with matched 5′ nucleotides. Genome Biol. 18 (1), 218.
49. Kulcsar, P.I. et al. (2017) Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. Genome Biol. 18, 190.
50. Chen, J.S. et al. (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550 (7676), 407-410.
51. Komor, A.C. et al. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533 (7603), 420-4.
52. Nishida, K. et al. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353 (6305), aaf8729.
53. Kim, Y.B. et al. (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol. 35 (4), 371-376.
54. Komor, A.C. et al. (2017) Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci Adv. 3 (8), eaao4774.
55. Rees, H.A. et al. (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun. 8, 15790.
56. Gaudelli, N.M. et al. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature 551 (7681), 464-471.
57. Komor, A.C. et al. (2017) Editing the Genome Without Double-Stranded DNA Breaks. ACS Chem Biol. (in press) DOI: 10.1021/acschembio.7b00710
58. Fu, Y. et al. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 32 (3), 279-284.
59. Chen, B. et al. (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155 (7), 1479-91.
60. Dang, Y. et al. (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol. 16, 280.
61. Konermann, S. et al. (2014) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583-588.
62. Xu, X. et al. (2016) A CRISPR-based approach for targeted DNA demethylation. Cell Discov. 2, 16009.
63. Zalatan, J.G. et al. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160 (1-2), 339-50.
64. Gao, Z. et al. (2017) Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. Transcription 8 (5), 275-287.
65. Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 32 (4), 347-55.
66. Nissim, L. et al. (2014) Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. Mol Cell. 54 (4), 698-710.

67. Gao, Y. and Zhao, Y. (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol. 56 (4), 343-9.
68. Yoshioka, S. et al. (2015) Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. Sci Rep. 5, 18341.
69. Xie, K. et al. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci U S A. 112 (11), 3570-5.
70. Port, F. and Bullock, S.L. (2016) Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cpf1 gRNAs. bioRxiv. DOI: 10.1101/046417
71. Qi, W. et al. (2016) High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. BMC Biotechnol. 16 (1), 58.
72. Schwartz, C.M. et al. (2016) Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. ACS Synth Biol. 5 (4), 356-9.
73. DeWitt, M.A. et al. (2017) Genome editing via delivery of Cas9 ribonucleoprotein. Methods 121-122, 9-15.
74. Kelley, M.L. et al. (2016) Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing. J Biotechnol. 233, 74-83.
75. Hendel, A. et al. (2015) Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nat Biotechnol. 33 (9), 985-989.
76. Rahdar, M. et al. (2015) Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. Proc Natl Acad Sci U S A. 112 (51), E7110-7.
77. Ryan, D.E. et al. (2018) Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. Nucleic Acids Res. 46 (2), 792-803.
78. Lee, K. et al. (2017) Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. Elife 6, e25312.
79. Holkers, M. et al. (2014) Adenoviral vector DNA for accurate genome editing with engineered nucleases. Nat Methods 11 (10), 1051-7.
80. Zhang, J.P. et al. (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol 18 (1), 35.
81. Maresca, M. et al. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. Genome Res. 23 (3), 539-46.
82. Yao, X. et al. (2017) CRISPR/Cas9 - Mediated Precise Targeted Integration In Vivo Using a Double Cut Donor with Short Homology Arms. EBioMedicine 20, 19-26.
83. He, X. et al. (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res. 44 (9), e85.
84. Suzuki, K. et al. (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature 540 (7631), 144-149.
85. Hisano, Y. et al. (2015) Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Sci Rep. 5, 8841.
86. Davis, L. and Maizels, N. (2014) Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. Proc Natl Acad Sci U S A. 111 (10), E924-32.
87. Gao, Y. et al. (2017) Single Cas9 nickase induced generation of NRAMP1 knockin cattle with reduced off-target effects. Genome Biol. 18 (1), 13.
88. Holliday, R. (2007) A mechanism for gene conversion in fungi. Genet Res. 89 (5-6), 285-307.
89. Richardson, C.D. et al. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol. 34 (3), 339-44.
90. Liang, X. et al. (2017) Enhanced CRISPR/Cas9-mediated precise genome editing by

improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. J Biotechnol. 241, 136-146.
91. Renaud, J.B. et al. (2016) Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. Cell Rep. 14 (9), 2263-2272.
92. Carlson-Stevermer, J. et al. (2017) Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. Nat Commun. 8 (1), 1711.
93. Ma, M. et al. (2017) Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. Cell Res. 27 (4), 578-581.
94. Carlson, D.F. et al. (2012) Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A. 109 (43), 17382-7.
95. Doyon, Y. et al. (2010) Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat Methods. 7 (6), 459-60.
96. Xiang, G. et al. (2017) Temperature effect on CRISPR-Cas9 mediated genome editing. J Genet Genomics. 44 (4), 199-205.
97. Duda, K. et al. (2014) High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. Nucleic Acids Res. 42 (10), e84.
98. Agudelo, D. et al. (2017) Marker-free coselection for CRISPR-driven genome editing in human cells. Nat Methods. 14 (6), 615-620.
99. Chu, V.T. et al. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 33 (5), 543-8.
100. Maruyama, T. et al. (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol. 33 (5), 538-42.
101. Ma, Y. et al. (2016) Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. RNA Biol. 13 (7), 605-12.
102. Robert, F. et al. (2015) Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. 7, 93.
103. Jayathilaka, K. et al. (2008) A chemical compound that stimulates the human homologous recombination protein RAD51. Proc Natl Acad Sci U S A. 105 (41), 15848-53.
104. Pinder, J. et al. (2015) Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. Nucleic Acids Res. 43 (19), 9379-92.
105. Song, J. et al. (2016) RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. Nat Commun. 7, 10548.
106. Yu, C. et al. (2015) Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell. 16 (2), 142-7.
107. Lin, S. et al. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3, e04766.

Chapter 2

In trans Paired Nicking Triggers **Seamless** Genome Editing **without** Double-stranded DNA Cutting.



Nat Commun. 2017 Sep 22;8(1):657.
Chen X, Janssen JM, Liu J, Maggio I, 't Jong AEJ, Mikkers HMM, Gonçalves MAFV.

Abstract

Precise genome editing involves homologous recombination between donor DNA and chromosomal sequences subjected to double-stranded DNA breaks made by programmable nucleases. Ideally, genome editing should be efficient, specific, and accurate. However, besides constituting potential translocation-initiating lesions, double-stranded DNA breaks (targeted or otherwise) are mostly repaired through unpredictable and mutagenic nonhomologous recombination processes. Here, we report that the coordinated formation of paired single-stranded DNA breaks, or nicks, at donor plasmids and chromosomal target sites by RNA-guided nucleases based on CRISPR-Cas9 components, triggers seamless homologydirected gene targeting of large genetic payloads in human cells, including pluripotent stem cells. Importantly, in addition to significantly reducing the mutagenicity of the genome modification procedure, this in trans paired nicking strategy achieves multiplexed, single-step, gene targeting, and yields higher frequencies of accurately edited cells when compared to the standard double-stranded DNA break-dependent approach.

Introduction

Programmable nucleases, and in particular RNA-guided nucleases (RGNs), are rendering genome editing applicable to numerous basic and applied research settings[1–3]. RGNs are ribonucleoprotein complexes formed by a guide RNA (gRNA) and a Cas9 protein with two nuclease domains, i.e., HNH and RuvC. RGNs cleave DNA complementary to the 5′ end of the gRNA when a contiguous protospacer adjacent motif (PAM) is present[3]. The fact that target DNA cutting is ultimately dictated by simple RNA-DNA hybridization rules confers versatility to RGN technologies[1–3]. A major drawback of conventional DNA editing stems, however, from the fact that double-stranded DNA break (DSB) repair in mammalian cells often takes place via mutagenic non-homologous end joining (NHEJ) instead of accurate homologous recombination (HR)[4]. As a result, allelic and non-allelic mutations, loss-of-heterozygosity, translocations, and other unwarranted genetic changes caused by on-target and off-target DSBs, are frequent[5]. Moreover, NHEJ also contributes to random and imprecise chromosomal insertion of the donor DNA[1, 6]. As a whole, these unpredictable genome-modifying events complicate the interpretation of experimental results and reduce the safety profile of candidate genetic therapies. Despite this, in certain experimental settings, such as those amenable to cell isolation and screening, homology-independent chromosomal DNA insertion is a valuable genetic modification strategy owing to its efficiency and applicability to non-dividing target cells[7–9].

Following from the above, developing new genome-editing principles that favor not only efficient but also precise homologydirected gene targeting in detriment of

mutagenic NHEJ are in demand. Indeed, emergent genome-editing research lines involve testing small RNAs, drugs, or viral proteins that steer DSB repair towards the HR pathway by inhibiting the competing NHEJ[10–12]. Parallel research lines exploit sequence-specific and strand-specific programmable nucleases ("nickases")[13–17] for generating single-stranded DNA breaks (SSBs), or nicks, which are noncanonical NHEJ substrates[4]. Besides bypassing DSB formation, "nickases" do not alter the regular cellular metabolism as small RNAs, drugs and viral proteins do. However, genome editing based on "nickases" is inefficient[13, 15–17]. In fact, the investigation of site-specific SSBs as triggers for homology-directed targeting of large DNA segments (e.g., entire transcriptional units) has not been explored.

Here, we investigate the feasibility of exploiting nicking RGNs containing the RuvC Cas9 mutant Asp10Ala ($Cas9^{D10A}$) or the HNH Cas9 mutant His840Ala ($Cas9^{H840A}$) to trigger genome editing via the simultaneous formation of SSBs at endogenous and exogenous DNA. We report that this strategy based on coordinated in trans paired nicking can improve the three main parameters of DNA editing, i.e., efficiency, specificity, and fidelity[1, 2] and achieves multiplexing homology-directed DNA addition of large genetic payloads.

Results

Mutagenesis caused by cleaving Cas9 vs. nicking Cas9

We started by confirming that unwarranted, potentially adverse, genome-modifying events (i.e., target allele mutagenesis and chromosomal translocations)[1] do occur more frequently in cells exposed to cleaving Cas9 than in those subjected to nicking Cas9 proteins. Firstly, we assessed the mutation rates resulting from RGN complexes consisting of cleaving (i.e., Cas9:$gRNA^{X}$) or nicking Cas9 nucleases (i.e., $Cas9^{D10A}$:$gRNA^{X}$ or $Cas9^{H840A}$:$gRNA^{X}$), where "X" symbolizes the target locus. The $Cas9^{D10A}$ and $Cas9^{H840A}$ proteins differ from wild-type Cas9 in that they have amino-acid substitutions disrupting the catalytic centers of their RuvC and HNH nuclease domains, respectively. As a result, RGN complexes with $Cas9^{D10A}$ and $Cas9^{H840A}$ induce sequence-specific and strand-specific breaks on opposite DNA chains, namely, on the chain complementary and non-complementary to the gRNA, respectively. The *AAVS1* locus at 19q13.42 was selected for these experiments owing to its frequent use as a "safe harbor" for the targeted chromosomal insertion of exogenous DNA[18]. This assessment is based on a series of studies showing that *AAVS1* integrants are neither disturbed by, nor disturb the surrounding genomic environment, providing for long-term and stable transgene expression in different cell types[18]. A target site genotyping assay in human embryonic kidney 293 T cells showed that Cas9:$gRNA^{S1}$ complexes targeting the *AAVS1* locus readily yielded substantially higher levels of DSBs than their $Cas9^{D10A}$:$gRNA^{S1}$ counterparts (**Supplementary Fig. 1a**). To augment the stringency of the genotyping assay, we next carried out dose–response

experiments in human cervix carcinoma HeLa cells using increasing amounts of adenoviral vectors encoding either Cas9 or Cas9^{D10A}, each mixed with a fixed amount of an adenoviral vector expressing a gRNA addressing each Cas9 protein to *AAVS1*. A direct relationship between the detection of small insertions and deletions (indels) and nuclease concentrations could be readily established after Cas9:gRNA^{S1} delivery, whereas this was much less so upon Cas9^{D10A}:gRNA^{S1} transfer (**Supplementary Fig. 1b**). These data directly correlated with the much higher frequencies of indel-derived *EGFP* disruption in EGFP+ H27 reporter cells triggered by cleaving Cas9:gRNA^{GFP2} when compared to those induced by nicking Cas9^{D10A}:gRNA^{GFP2} or by Cas9^{H840A}:gRNA^{GFP2} complexes (**Supplementary Fig. 1c**).

Secondly, we setup a PCR assay to compare the assembly of chromosomal translocations caused by the formation of DSBs vs. SSBs at two distinct loci. To this end, HeLa cells were transfected with plasmids coding for cleaving or nicking RGNs targeting *DMD* and *AAVS1* sequences. Amplicons diagnostic for translocation events between *DMD* and *AAVS1* were exclusively detected in cells exposed to the cleaving RGNs (**Supplementary Fig. 1d**). Sanger sequencing of individual amplicons established their origin at t(*X*;19)(p21;q13) (**Supplementary Fig. 1e**). Taken together, these experiments formally demonstrate that unwarranted, potentially adverse, genome-modifying events occur more frequently in cells receiving RGNs containing cleaving Cas9 than in those harboring nicking Cas9^{D10A}.

In trans paired nicking yields seamless *DMD* gene targeting

Next, we sought to investigate homology-directed gene targeting based on inducing DSBs vs. SSBs not only at acceptor chromosomal sequences but also at donor DNA templates. The *DMD* gene at Xp21.2 was chosen as target locus. By spanning over 2.4 Mb, *DMD* is the largest human protein-coding gene known. Of note, defective *DMD* alleles cause Duchenne muscular dystrophy (DMD), a progressive lethal neuromuscular disease affecting ~1 in 3500–5000 boys[19, 20]. For these experiments, we generated plasmid pgRNA^{DMD}, to address Cas9 proteins to *DMD* intron 43, and EGFP-encoding constructs pDonor^{DMD} and pDonor$^{DMD.TS}$ to serve as exogenous HR substrates (**Fig. 1a**). Construct pDonor$^{DMD.TS}$ differs from pDonor^{DMD} in that it has a target site (TS) for gRNA^{DMD} next to its targeting module (**Fig. 1a**). Importantly, all transgene-containing donors used in the present study have autonomous transcription units, which in contrast to splice acceptor-containing gene trapping constructions, avoid biased selection of on-target integrants[21]. Genome-editing experiments were initiated by exposing HeLa cells to pDonor^{DMD} and cleaving Cas9:gRNA^{DMD} complexes (standard setting) or to pDonor$^{DMD.TS}$ and nicking Cas9^{D10A}:gRNA^{DMD} complexes (in trans paired nicking; Nick^2). After eliminating episomal DNA by sub-culturing, genetically modified cells were quantified through flow cytometry. This analysis revealed that the in trans paired nicking strategy led to significantly

higher percentages of genetically modified cells when compared to those obtained through the standard approach (**Fig. 1b**). Similar results were obtained by using donor constructs whose *DMD*-targeting modules were flanked by the gRNA^{DMD} TS in a direct or inverted repeat orientation (**Supplementary Fig. 2**). These data are consistent with earlier theoretical models and more recent experimental systems indicating a role for nicked HR partners as recombination-initiating substrates[22, 23]. Of note, although at this target sequence paired DSB formation (in trans paired breaking; DSB^2) yielded the highest frequencies of EGFP+ cells, the resulting free-ended HR substrates are prone to aberrant concatemer assembly (see below). Indeed, it has been previously shown that the DSB^2 strategy results in higher frequencies of random chromosomal insertions through illegitimate recombination processes when compared to those obtained by the standard DSB-dependent gene targeting approach[6]. Conversely, consistent with previous studies[13, 15–17], generating SSBs exclusively at chromosomal DNA yielded the lowest frequencies of stably transfected cells.

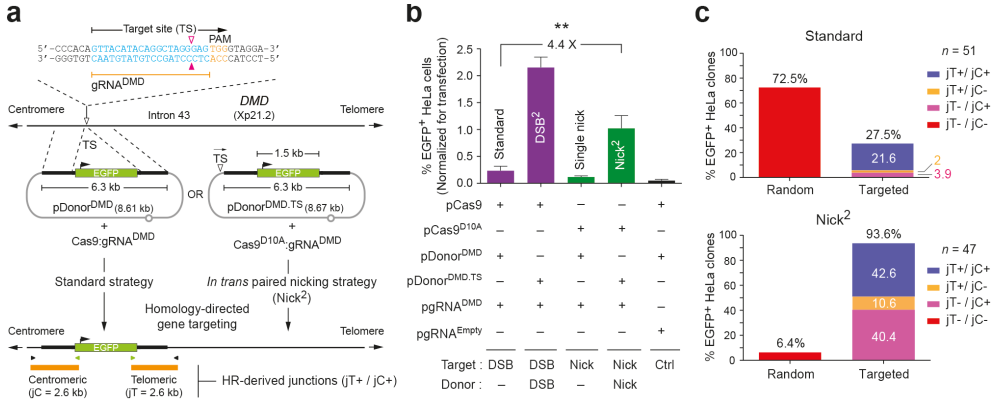


Fig. 1
Homology-directed *DMD*-targeting using standard and in trans paired nicking strategies. **a** Schematics of standard and in trans paired nicking (Nick^2) procedures. The former involve DSB formation only at the target sequence; the latter comprise SSB formation at target plus donor sequences. pDonor^{DMD} and pDonor$^{DMD.TS}$ have their transgenes flanked by sequences identical to those framing the gRNA^{DMD} target site (TS). *Open and solid magenta arrowheads*, position of the phosphodiester bond cleavage induced by Cas9's RuvC and HNH nuclease domains, respectively. *Solid arrowhead*, position of the SSB induced by Cas9^{D10A}. The modified pDonor$^{DMD.TS}$ differs from pDonor^{DMD} in that it has the gRNA^{DMD} TS next to its targeting module. The transgene is formed by human *PGK1* promoter, *EGFP* ORF, and bovine *GH1* polyadenylation signal sequences. Cas9:gRNA^{DMD} and Cas9^{D10A}:gRNA^{DMD} are cleaving and nicking RGN complexes, respectively. PAM protospacer adjacent motif. An integrant generated by HR events at both ends of the targeting module is depicted. The amplicons specific for telomere-sided and centromere-sided transgenic-*DMD* junctions (jT and jC, respectively), are equally shown. *Horizontal arrowheads*, primers. **b** Quantification of stable transfection levels by flow cytometry. Flow cytometry of long-term HeLa cell cultures initially exposed to the indicated plasmids. The *bars* correspond to mean ± s.d. of three independent biological replicates done on different days. $^{**}P = 0.006$ (two-tailed *t*-test). **c** Cumulative molecular characterization of integrants generated by the conventional vs. the in trans paired nicking strategies. The frequencies of clones with random insertions (jT−/jC−), HR-derived telo-

meric junctions (jT+/jC−), HR-derived centromeric junctions (jT−/jC+) and HR-derived telomeric and centromeric junctions (jT+/jC+) are plotted. The corresponding PCR screening data are presented in **Supplementary Fig. 3**

Subsequently, we compared in trans paired nicking with standard gene targeting in terms of their relative specificities and fidelities. The specificity is ascertained by detecting donor sequences at the target site; the fidelity is established by demonstrating that telomere-sided and centromere-sided junctions between donor and target DNA are formed through error-free HR (jT+ and jC+, respectively). Randomly selected EGFP+ HeLa clones (n = 98) were screened via PCR assays targeting both junctions (**Fig. 1c** and **Supplementary Fig. 3**). In the set of clones modified through the delivery of pDonor^{DMD}, pCas9 and pgRNA^{DMD} (n = 51), the *DMD*-targeted fraction was 27.5% with 21.6% of these integrants being accurately targeted (jT+/jC+). Notably, in the set of clones modified via the transfer of pDonor$^{DMD.TS}$, pCas9^{D10A} and pgRNA^{DMD} (n = 47), these fractions were 93.6% and 42.6%, respectively (**Fig. 1c**). We conclude that, when compared to conventional DSB-induced gene targeting, in trans paired nicking was more efficient, specific, and accurate at the *DMD* locus.

In trans paired nicking yields seamless *AAVS1* gene targeting

We next examined the performance of in trans paired nicking and standard gene targeting at *AAVS1* (**Fig. 2a**). As aforementioned, this locus is commonly used as a "safe harbor" for the chromosomal insertion of exogenous DNA in human cells[18]. These experiments were initiated by transfecting HeLa and 293 T cells with pDonor.E^{S1} or pDonor.E$^{S1.TS}$ each mixed with plasmids encoding either Cas9:gRNA^{S1} or Cas9^{D10A}:gRNA^{S1} (**Fig. 2a**). The pDonor.E$^{S1.TS}$ construct has its targeting module flanked by two gRNA^{S1} TS (**Fig. 2a**). The rationale for this donor design was provided by the experiments showing that such arrangement yields significantly higher frequencies of stably transfected cells when compared to isogenic templates containing a single gRNA^{S1} TS (**Supplementary Fig. 4**). In agreement with the *DMD*-targeting experiments, when compared to experiments involving single DSBs (standard setting) or single SSBs, in trans paired nicking of *AAVS1* and pDonor.E$^{S1.TS}$ led to significantly higher percentages of genetically modified cells (**Fig. 2b**). Similar results were gathered by using different gRNA and donor DNA reagents or the alternative nicking Cas9^{H840A} variant whose inactivated HNH domain assures that SSBs occur at the DNA chain opposite to that hydrolyzed by its RuvC-disabled Cas9^{D10A} counterpart (**Supplementary Fig. 5**). Importantly, amplicons diagnostic for HR-derived integrants were readily retrieved not only from cells subjected to inaccurate DNA editing by paired DSB formation but also from cells exposed to the accurate in trans paired nicking procedure (**Fig. 2c**). Indeed, in striking contrast to inducing in trans paired DSBs (DSB^2), generating in trans paired SSBs (Nick^2), did not result in the assembly of disruptive donor DNA concatemers (**Fig. 2c**), presumably emerging through ligation of free-ended termini generated *in cellula* by Cas9:gRNA^{S1} [6]. Final-

ly, we probed an alternative in trans paired nicking gene targeting strategy in which two different gRNAs generate tandem SSBs within the interacting homologous sequences. This strategy, tandem paired nicking, yielded stable transfection levels that

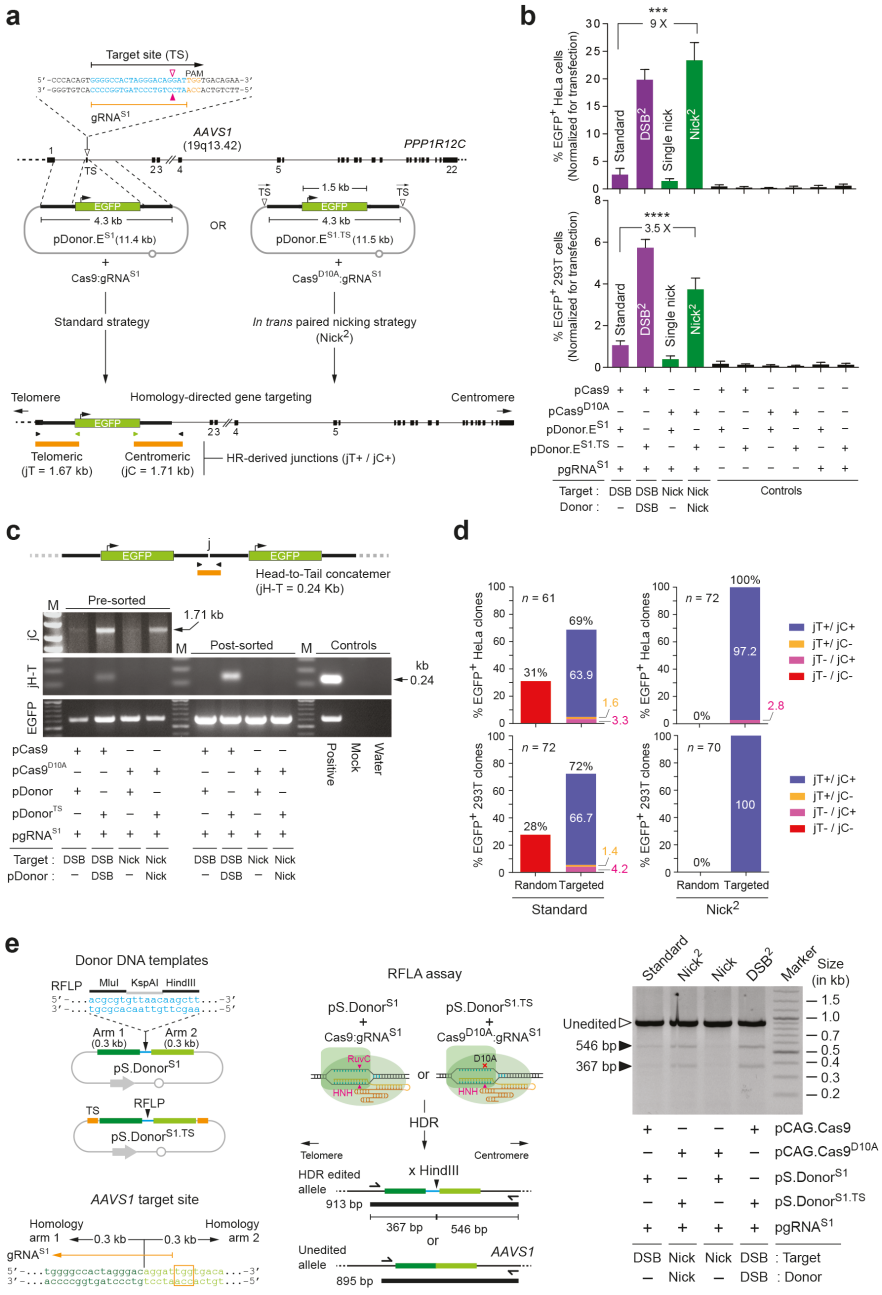


Fig. 2
Homology-directed *AAVS1* targeting using standard and in trans paired nicking strategies.
a Diagram of standard and in trans paired nicking (Nick²) procedures. The former involve DSB formation only at the target sequence; the latter comprise SSB formation at target

plus donor sequences. pDonor^{S1} and pDonor$^{S1.TS}$ have their transgenes framed by sequences homologous to *AAVS1*. pDonor$^{S1.TS}$ differs from pDonor^{S1} in that it has the gRNA^{S1} target site (TS) bracketing its EGFP-encoding targeting module. Cas9:gRNAS1 and Cas9^{D10A}:gRNA^{S1} are cleaving and nicking RGNs, respectively. *Open and solid magenta arrowheads*, position of the phosphodiester bond cleavage induced by Cas9's RuvC and HNH nuclease domains, respectively. *Solid arrowhead*, position of the SSB induced by Cas9^{D10A}. Amplicons diagnostic for telomere-sided and centromere-sided transgenic-*AAVS1* junctions (jT and jC, respectively), are depicted. **b** Quantification of stably transfected cells. Flow cytometry of long-term HeLa and 293T cell cultures initially transfected with the indicated plasmids. The bars correspond to mean ± s.d. of six biological replicates from two independent experiments (three biological replicates per experiment). ****$P < 0.0001$ (two-tailed *t*-tests). **c** Probing for wanted (gene targeting) and unwanted (concatemerization) genome-modifying events. Amplicons diagnostic for gene targeting (jC) and head-to-tail concatemers (jH-T) in 293T cell populations transfected with the indicated constructs are presented. This assay was also run on EGFP-sorted cells (post-sorted). *EGFP* served as an internal control template. **d** Cumulative molecular characterization of integrants generated by the conventional and in trans paired nicking strategies. The frequencies of 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 respective PCR screening data are presented in **Supplementary Fig. 7**. **e** Homology-directed *AAVS1* editing after inducing DSBs or SSBs. pS.Donor^{S1} and pS.Donor$^{S1.TS}$ have a restriction-fragment length polymorphism (RFLP) flanked by 300-bp *AAVS1* sequences ("arms"). pS.Donor$^{S1.TS}$ has the gRNA^{S1} TS flanking its targeting module (*orange boxes*). RFLA restriction-fragment length analysis; *half arrows* primers; PAM boxed sequence. RFLA products diagnostic for unedited and HR-edited *AAVS1* alleles retrieved from HeLa cells transfected with the indicated plasmid combinations are identified by *open and closed arrowheads*, respectively

were within the range of those achieved by using the standard, DSB-dependent, gene targeting procedure (**Supplementary Fig. 6**).

To gauge the specificity and fidelity resulting from in trans paired nicking vs. standard gene targeting at *AAVS1*, randomly selected EGFP+ clones (n = 275) were isolated from HeLa and 293 T cell populations and were screened through junction PCR (**Fig. 2d** and **Supplementary Fig. 7**). We observed that 63.9% and 66.7% of the HeLa and 293 T cells exposed to the standard setting underwent accurate homology-directed gene targeting (jT+/jC+), respectively (**Fig. 2d**). In the remaining clones, illegitimate recombination led instead to off-target integrants (jT−/jC−) and to on-target integrants lacking HR-derived junctions either from the centromeric or telomeric side (jT+/jC− or jT−/jC+, respectively). Remarkably, the fraction of properly targeted HeLa and 293 T cells subjected to in trans paired nicking was as high as 97.2 and 100%, respectively (**Fig. 2d**). Finally, Sanger sequencing established that precisely targeted integrants resulting from in trans paired nicking and conventional gene targeting were undistinguishable (**Supplementary Fig. 8**).

To complement the previous gene targeting experiments involving sizable and transcriptionally active donor constructs, we next asked whether short, transcriptionally inert donor constructs, can equally serve as in trans paired nicking substrates. To this end, *AAVS1*-targeting plasmids pS.Donor^{S1} and pS.Donor$^{S1.TS}$, resistant and susceptible to RGNs, respectively (**Fig. 2e**, *left panel*), were transfected into human cells together with constructs expressing Cas9:gRNA^{S1} or Cas9^{D10A}:gRNA^{S1} (**Fig. 2e**, *mid-*

dle panel). HR engaging pS.Donor^{S1} or pS.Donor$^{S1.TS}$ sequences should result in the targeted chromosomal insertion of 18-bp DNA fragments incorporating restriction enzyme polymorphisms (**Fig. 2e**, *middle panel*). Detection of these genome-editing events by restriction enzyme fragment length analysis (RFLA) revealed that in trans paired nicking is compatible with the use of short, transcriptionally inert, donor DNA templates (**Fig. 2e**, *right panel*).

Paired RGNs inducing offset nicks on opposite chromosomal DNA strands ensure that DSBs are mostly restricted to their bipartite target sequences owing to the coordinated and local formation of SSBs on both polynucleotide chains[24, 25]. The resulting gains in DNA cutting specificity render this dual RGN approach appealing, hereafter named in cis paired nicking for the sake of consistency. Hence, albeit dependent on two gRNAs and on the generation of mutagenic DSBs, we sought nonetheless to compare in cis with in trans paired nicking as stimuli for site-specific chromosomal DNA insertion (knock-in). Therefore, in addition to the four experimental conditions tested before (**Fig. 2b**), in these new experiments, we transfected human cells with pDonor^{S1} and pCAG.Cas9^{D10A} mixed with constructs expressing two different *AAVS1*-specific gRNA pairs (i.e., gRNA^{S1}/gRNA$^{S1.2}$ or gRNA^{S1}/gRNA$^{S1.3}$). Consistent with the previous data (**Fig. 2b**), the in trans paired nicking setup yielded the highest frequencies of genetically modified cells. The in cis paired nicking strategy led, in turn, to frequencies of genetically modified cells that were in the range of those obtained by inducing DSBs or SSBs exclusively at the target site (**Supplementary Fig. 9**).

In trans paired nicking in pluripotent stem cells

Despite their patent scientific and biomedical importance, genetic manipulation of human pluripotent stem cells (PSCs) remains limited by the typically low efficiency, specificity, and accuracy of homology-directed gene targeting, even when using programmable nucleases (see e.g., ref. [21]). Therefore, we investigated the performance of in trans paired nicking in human induced PSCs (iPSCs; **Supplementary Fig. 10**) and human embryonic stem cells (ESCs)[26]. In addition to pDonor^{S1} and pDonor$^{S1.TS}$ (**Supplementary Fig. 4a**), we included in these experiments, pDonor.EP^{S1} and pDonor.EP$^{S1.TS}$ encoding Puro^{R}.T2A.EGFP instead of EGFP. The data generated with these new HR substrates in HeLa cells (**Supplementary Fig. 11**) were similar to those of previous experiments showing the superiority of in trans paired nicking over standard gene targeting in achieving efficient cell engineering at *AAVS1* (**Fig. 2b** and **Supplementary Figs. 4b–6** and **9**). Importantly, this superiority was equally established in iPSCs and ESCs by using dual-color flow cytometry and colony-formation assays involving the detection of EGFP+/TRA-1-81+ cells (**Fig. 3a**, **b**) and puromycin-resistant colonies stained for alkaline phosphatase, respectively (**Fig. 3c**). In addition, when compared to in trans paired nicking, DSB-triggered *AAVS1* targeting induced

higher frequencies of apoptotic Annexin V+ cells in ESC cultures (**Supplementary Fig. 12**). These results are consistent with the well-established sensitivity of PSCs to DSBs[27].

To determine the precision of genome editing in iPSCs subjected to in trans paired nicking vs. standard genome-editing protocols, puromycin-resistant clones (n = 80)

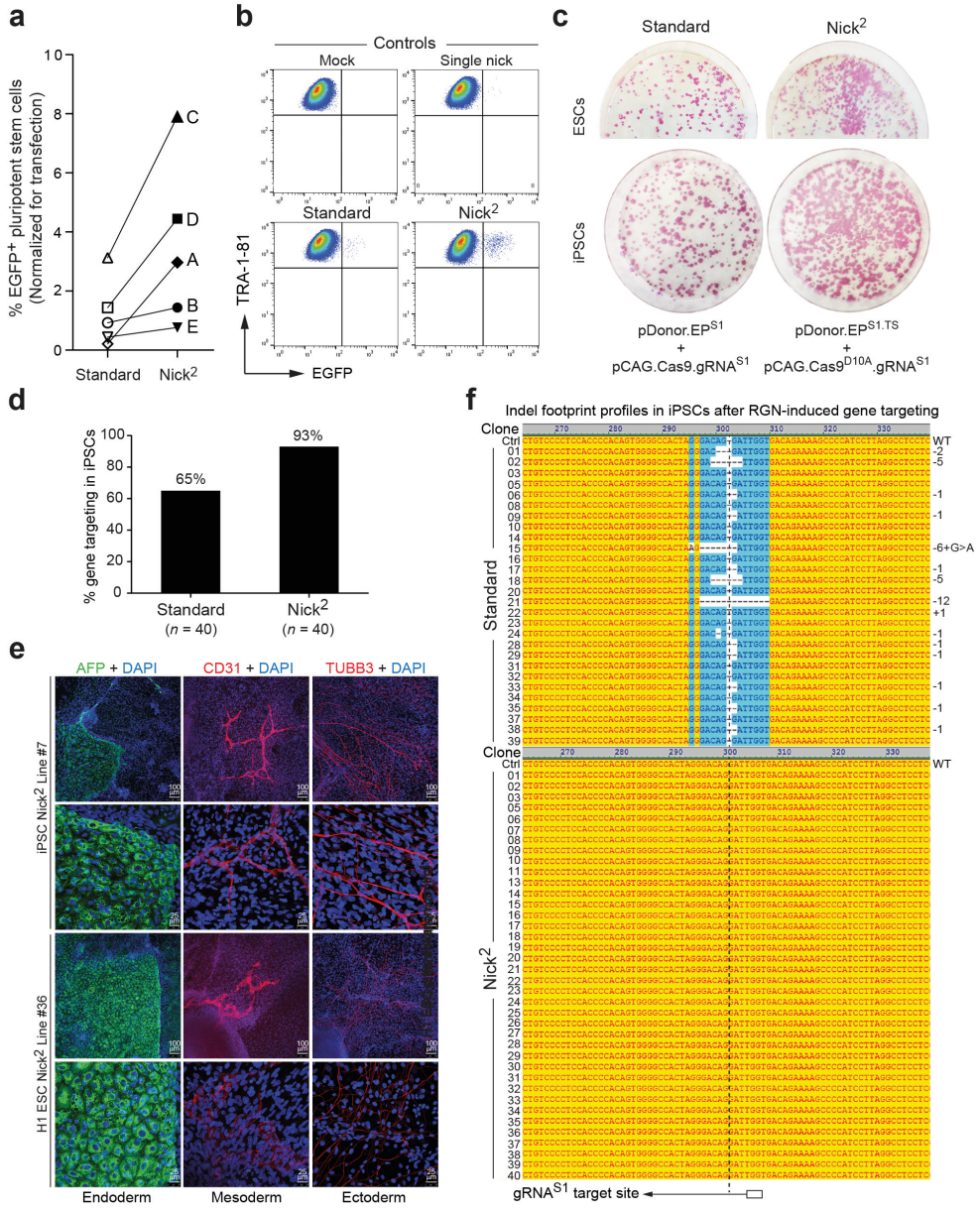


Fig. 3
Comparing RGN-induced gene targeting based on standard and in trans paired nicking in human PSCs. **a** Quantification of genetically modified PSCs by flow cytometry. Cultures of iPSCs (A, B, and E) and ESCs (C and D) were exposed to *AAVS1*-specific cleaving Cas9:gRNA^{S1} (standard) or nicking Cas9^{D10A}:gRNA^{S1} (Nick^{2}) complexes mixed with RGN-resistant or RGN-susceptible donor constructs, respectively, encoding either EGFP or Puro^{R}.T2A.EGFP. The frequencies of gene-modified PSCs were determined by flow cytometric quantification of EGFP+ and TRA-1-81+ dually labeled cells. **b** Representative flow cytometry dot plots corresponding to RGN-induced gene targeting experiments in PSCs. **c** Detection of gene-modified PSCs by colony-formation assays. ESCs (*top*) and iPSCs (*bottom*) were co-transfected with the indicated plasmids. After puromycin selection, alkaline phosphatase staining identified genetically modified PSC colonies. **d** RGN-induced gene targeting frequencies at *AAVS1* in iPSCs. Junction PCR analyses of puromycin-resistant colonies from iPSC cultures initially co-transfected with pDonor.EP^{S1} and pCas9.gRNA^{S1} (standard) or with pDonor.EP^{S1}.TS and pCas9^{D10A}.gRNA^{S1} (Nick2). The respective PCR screening data are presented in **Supplementary Fig. 13**. **e** Differentiation potential of gene-edited PSCs. ESC and iPSC lines were targeted at *AAVS1* by in trans paired nicking. Cell types characteristic of ectoderm, endoderm, and mesoderm were identified by confocal immunofluorescence microscopy for TUBB3, AFP, and CD31, respectively. **f** Characterization of indel footprints in iPSCs subjected to standard vs. in trans paired nicking. Nucleotide sequencing of *AAVS1* target alleles in randomly selected iPSC clones (n = 68) genetically modified by DSB-dependent and in trans paired nicking methodologies (Standard and Nick2, respectively). Indel footprints were exclusively identified in iPSCs subjected to the standard gene targeting approach (15/28). The gRNA^{S1} target site is indicated underneath the sequence reads. *Open box* PAM; *vertical dashed line* position of expected RGN-induced phosphodiester bond cleavage; *Ctrl* reference wild-type nucleotide sequence from unedited cells

were screened with a PCR assay specific for HR-derived junctions (**Fig. 3d** and **Supplementary Fig. 13**). The gene targeting specificity in iPSCs exposed to standard and in trans paired nicking procedures was 65 and 93%, respectively (**Fig. 3d** and **Supplementary Fig. 13**). Contributing to the difficulty in isolating iPSC lines that undergo seamless genome editing is the fact that a sizable fraction of cells, in addition to the intended genetic modification at one of the target alleles, harbor mutations at the other allele[28]. These mutations correspond to unpredictable indel footprints created after NHEJ-mediated repair of targeted DSBs[28]. Hence, to further characterize the genetically modified iPSCs, nucleotide sequence analysis of target DNA was performed in individual iPSC clones subjected to standard and in trans paired nicking protocols. This analysis revealed the presence of a range of indel footprints exclusively in the iPSC lines generated by standard gene targeting (**Fig. 3f**). Indeed, the *AAVS1* target site remained pristine in all of the randomly selected iPSC lines obtained after applying the in trans paired nicking protocols (**Fig. 3f**). These results are in agreement with our previous data (**Supplementary Fig. 1**) and the fact that, in contrast to DSBs, SSBs are not canonical substrates for NHEJ.

Finally, iPSC lines genetically engineered through standard and in trans paired nicking remained pluripotent (**Fig. 3e** and **Supplementary Fig. 14**). We conclude that, instead of generating DSBs, targeted DNA integration at the *AAVS1* "safe harbor" in different cell types is best achieved via coordinated RGN-induced paired nicking of donor and acceptor DNA.

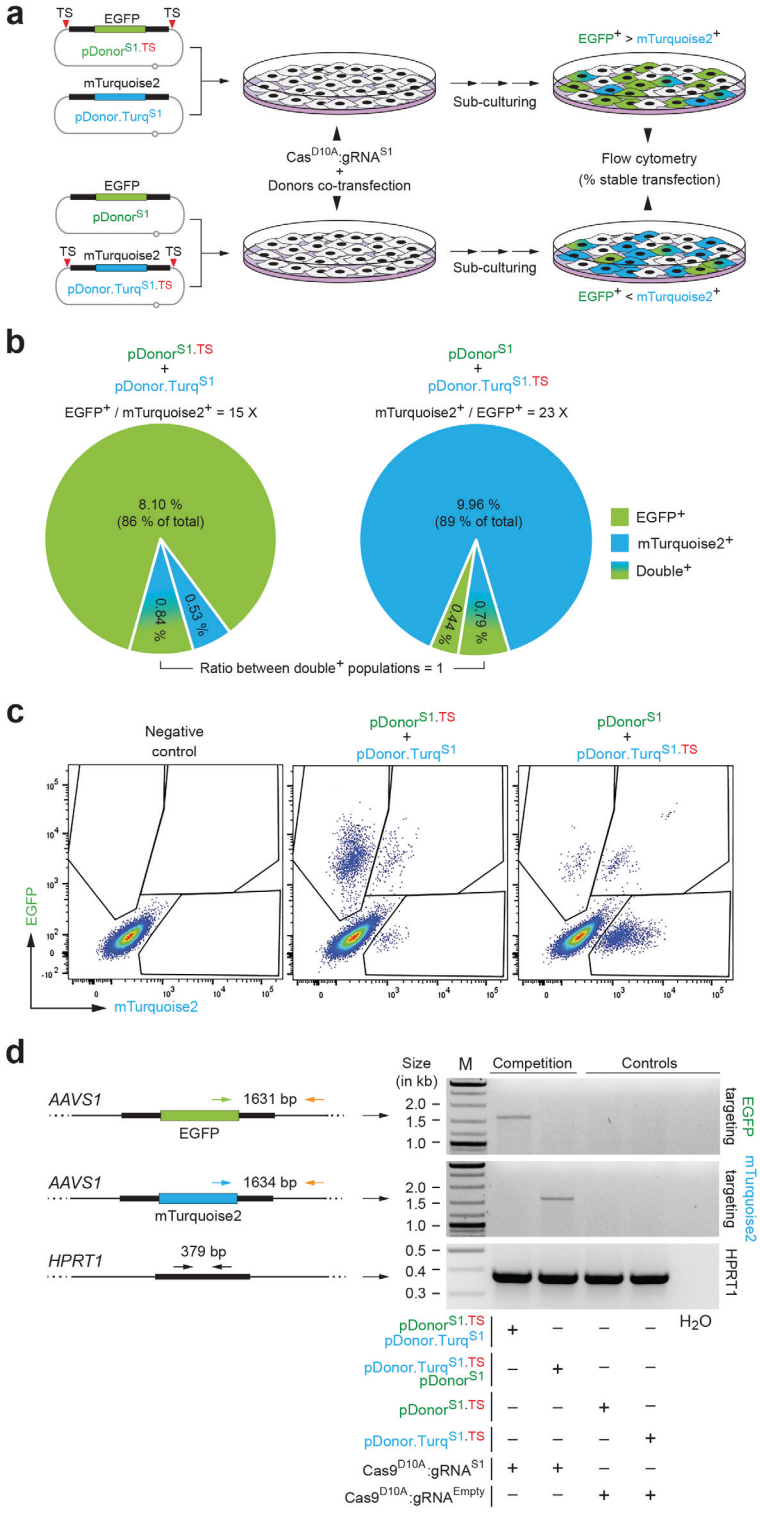

a
TS
EGFP
TS
pDonorS1.TS
mTurquoise2
pDonor.TurqS1
CasD10A:gRNAS1
+
Donors co-transfection
Sub-culturing
EGFP+ > mTurquoise2+
Flow cytometry
(% stable transfection)
EGFP
pDonorS1
TS
mTurquoise2
TS
pDonor.TurqS1.TS
Sub-culturing
EGFP+ < mTurquoise2+
b
pDonorS1.TS
+
pDonor.TurqS1
EGFP+ / mTurquoise2+ = 15 X
8.10 %
(86 % of total)
0.84 %
0.53 %
pDonorS1
+
pDonor.TurqS1.TS
mTurquoise2+ / EGFP+ = 23 X
9.96 %
(89 % of total)
0.44 %
0.79 %
EGFP+
mTurquoise2+
Double+
Ratio between double+ populations = 1
c
Negative
control
pDonorS1.TS
+
pDonor.TurqS1
pDonorS1
+
pDonor.TurqS1.TS
EGFP
mTurquoise2
d
AAVS1
1631 bp
EGFP
AAVS1
1634 bp
mTurquoise2
HPRT1
379 bp
Size
(in kb)
M
Competition
Controls
2.0
1.5
1.0
2.0
1.5
1.0
0.5
0.4
0.3
EGFP targeting
mTurquoise2 targeting
HPRT1
H2O
pDonorS1.TS
pDonor.TurqS1
+ – – –
pDonor.TurqS1.TS
pDonorS1
– + – –
pDonorS1.TS
– – + –
pDonor.TurqS1.TS
– – – +
Cas9D10A:gRNAS1
+ + – –
Cas9D10A:gRNAEmpty
– – + +

Fig. 4 Competition for gene targeting between donor DNA resistant and sensitive to RGN-induced nicking. **a** Schematics of the experimental design. HeLa cells were co-transfected with the indicated donor templates together with plasmids encoding nicking Cas9^{D10A}:gRNA^{S1}. **b** Quantification of stably transfected cell populations. The frequencies of genetically modified cells were determined at 27 days post-transfection by EGFP-directed and mTurquoise2-directed flow cytometry. The ratios between the frequencies of the various gene-modified subpopulations are presented. **c** Flow cytometry dot plots corresponding to the end-point of the experiments. Mock-transfected cultures served to set the thresholds for background fluorescence (negative control). **d** Gene targeting in cells containing donor DNA resistant and susceptible to RGN nicking. Amplicons diagnostic for homology-directed gene targeting involving EGFP-encoding and mTurquoise2-encoding donor templates are indicated. *HPRT1* provided for an internal control target sequence

Multiplexing gene targeting by in trans paired nicking

To confirm that *AAVS1*-targeting donor DNA subjected to RGN nicking is a superior substrate for site-specific chromosomal DNA insertion, we setup competition experiments involving the co-targeting of two donors each encoding a different reporter, i.e., EGFP or mTurquoise2 (**Fig. 4a**). For these experiments, one of the two donors contained TS sequences, whereas the other did not (**Fig. 4a**). Flow cytometry showed that pDonor$^{S1.TS}$ and pDonor.Turq$^{S1.TS}$ subjected to RGN-induced nicking led to 15-fold and 23-fold higher frequencies of genetically modified cells, respectively, when compared to their competitor, RGN-resistant, donor counterparts pDonor^{S1} and pDonor.Turq^{S1} (**Fig. 4b**, **c**). Consistent with these results, homology-directed gene targeting in cells containing both RGN-resistant and RGN-susceptible donors involved primarily the latter substrates, independently of the product that they encoded (**Fig. 4d**).

Hitherto, multiplexing genome editing has primarily entailed NHEJ-based manipulations such as those involving RGN pairs for knocking-out two genes simultaneously or for creating chromosomal deletions[1]. Such approaches are, however, not applicable for the targeted addition of new genetic information. For this purpose, multiplexing homology-directed DNA insertion based on different donor constructs can, in principle, be used instead. Unfortunately, HR-dependent chromosomal knock-in of two different donors in individual cells is a very rare event. Moreover, in addition to generating high frequencies of indel footprints, the necessary programmable nuclease pairs can induce loss-of-heterozygosity and/or translocations (**Supplementary Fig. 1**). Therefore, engineering cells with exogenous DNA inserted at two different loci or at two alleles of a single locus (bi-allelic targeting) is normally a complex and time-consuming procedure. Indeed, these procedures include constructing donors with positive/negative selection markers for isolating and screening the few cells that undergo seamless gene targeting, often followed by marker removal. This lengthy process is subsequently repeated on the selected cell clone(s) using, this time, a second donor construct.

We thus sought to capitalize on the higher efficiency, specificity and accuracy of

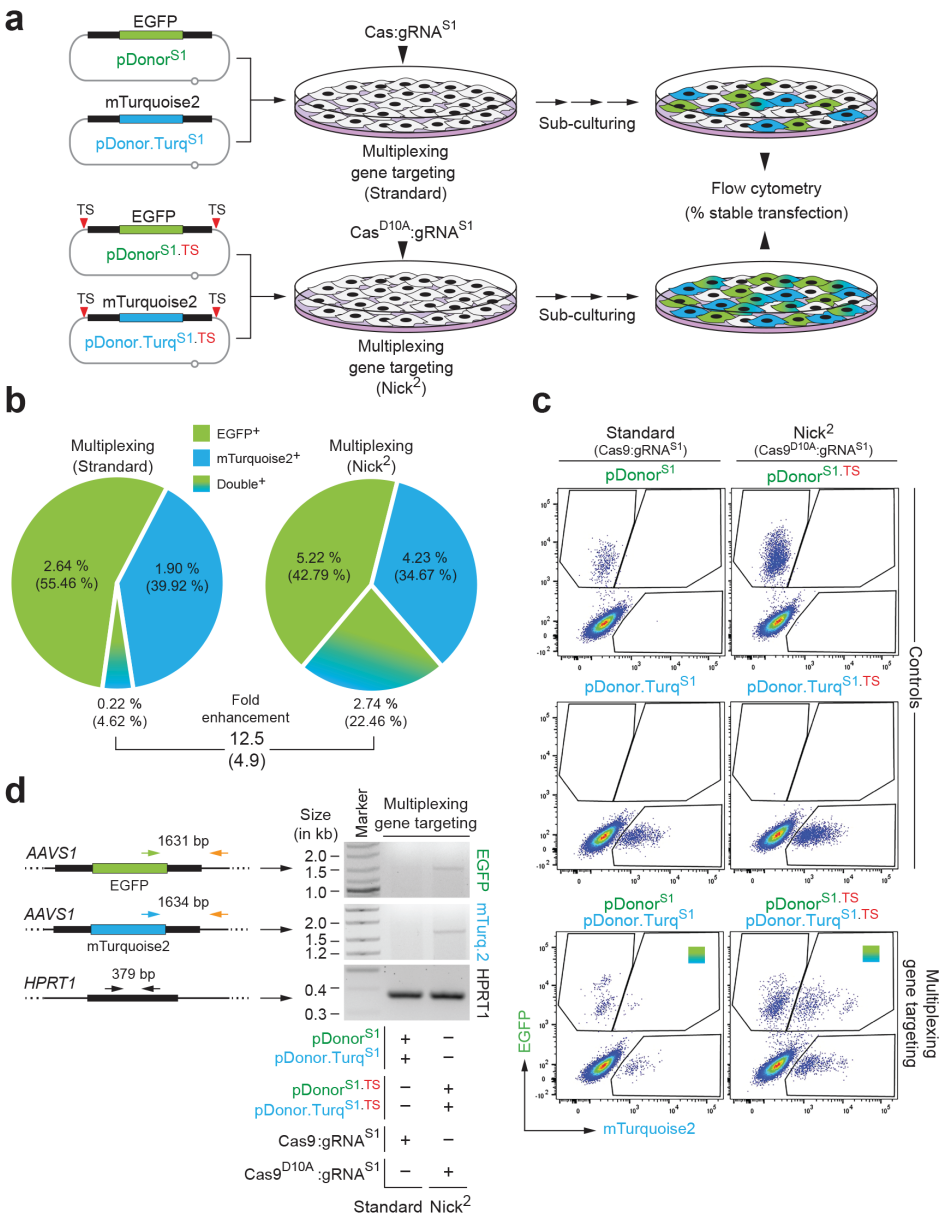


Fig. 5 Multiplexing homology-directed DNA addition. **a** Diagram of the experimental design. HeLa cells were co-transfected with the indicated donor constructs together with plasmids encoding either cleaving Cas9:gRNA^{S1} or nicking Cas9^{D10A}:gRNA^{S1} complexes. **b** Quantification of stably transfected cell populations. The frequencies of genetically modified cells were determined at 27 days post-transfection by EGFP-directed and mTurquoise2-directed flow cytometry. The ratios between the frequencies of the double-positive cell populations generated by standard and in trans paired nicking multiplexing, are presented. *Numerals* between *brackets* correspond to the fraction of each gene-modified subpopulation. **c** Flow cytometry dot plots corresponding to the end-point of the experiments. Parallel cultures transfected with a single donor construct mixed with plasmids expressing Cas9:gRNA^{S1} or

Cas9^{D10A}:gRNA^{S1} served as controls for setting the thresholds for EGFP and mTurquoise2 detection. **d** Gene co-targeting in cells containing a mixture of two donors resistant or susceptible to RGN nicking. PCR products specific for homology-directed gene targeting involving EGFP-encoding and mTurquoise2-encoding donor templates are indicated. *HPRT1* provided for an internal control target sequence

in trans paired nicking over the conventional DSB-dependent strategy at *AAVS1*, for testing one-step co-targeting of different alleles. These multiplexing knock-in experiments were initiated by exposing HeLa cells to pDonor$^{S1.TS}$, pDonor.Turq$^{S1.TS}$, and nicking Cas9^{D10A}:gRNA^{S1} (**Fig. 5a**). Controls consisted of treating HeLa cells with pDonor^{S1}, pDonor.Turq^{S1}, and cleaving Cas9:gRNA^{S1} (**Fig. 5a**). Remarkably, in comparison with the control setting, the multiplexing approach based on in trans paired nicking yielded one order of magnitude higher amounts of doubly-labeled EGFP+/mTurquoise2+ cells as measured by flow cytometry (**Fig. 5b**, **c**). These results directly correlated with the detection of HR-specific amplicons in parallel genomic DNA samples (**Fig. 5d**). After flow cytometry-assisted sorting of these EGFP+/mTurquoise2+ cells (**Supplementary Fig. 15**), single-cell clonal analysis (n = 35) revealed that 89% of them underwent *AAVS1*-targeting events, of which 94% were bi-allelic events involving both donor DNA templates (**Supplementary Fig. 16a**, **b**). An independent assay based on Southern blot analysis confirmed co-targeting of both expression units in individual cells without evidence for random chromosomal DNA insertion (**Supplementary Fig. 16a**, **c**). Taken together, these data show that simultaneous in trans paired nicking of independent donor substrates can provide for a simpler and faster strategy for achieving, in a seamless manner, multiplexed addition of foreign DNA into the genome of human cells.

In trans paired nicking yields seamless gene editing at *CCR5*

The product of the C–C motif chemokine receptor 5 gene *CCR5*, located at 3p21.31, serves as an HIV-1 co-receptor on macrophages and T cells[29]. Crucially, individuals homozygous for a 32-bp deletion disrupting CCR5 function (*CCR5Δ32*) are healthy and refractory to R5-tropic HIV-1 infection[29]. Hence, this locus is an appealing target for testing HIV therapies based on viral co-receptor knockout and site-specific "stacking" of restriction factor genes[29]. In addition, similarly to *AAVS1*, *CCR5* is frequently used as a generic "safe harbor" for the targeted chromosomal insertion of foreign DNA in human cells[18]. Thus, we next sought to compare DSB-dependent vs. SSB-dependent genome-editing approaches at *CCR5* after delivering RGNs together with *CCR5*-targeting constructs pS.Donor^{R5} or pS.Donor$^{R5.TS}$ marked with restriction enzyme polymorphisms (**Fig. 6a**). In these experiments, RFLA and mismatch-sensing T7 endonuclease I (T7EI) genotyping assays were deployed for assessing genomic changes through HR and/or NHEJ (**Fig. 6b**). Human cells treated with in trans paired nicking (Nick^2) and in trans paired breaking (DSB^2) protocols readily yielded noticeable HR-specific RFLA products (**Fig. 6c**, *top panel*). A prepon-

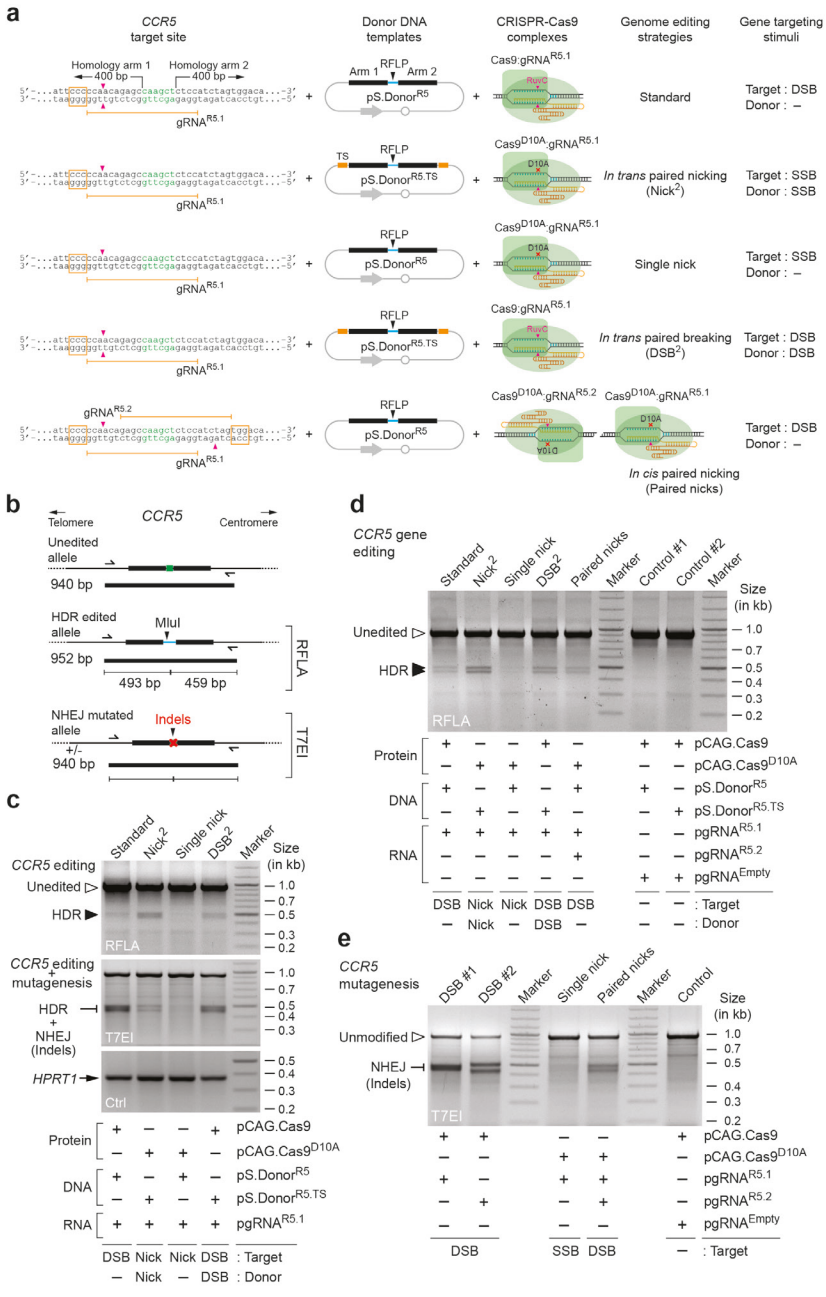


Fig. 6
Homology-directed *CCR5* editing after DSB vs. SSB generation. **a** Diagram of the different DSB-dependent and SSB-dependent genome-editing strategies. pS.Donor^{R5} and pS.Donor$^{R5.TS}$ have a restriction-fragment length polymorphism (RFLP) flanked by 400-bp *CCR5* sequences ("arms"). pS.Donor$^{R5.TS}$ has the gRNA$^{R5.1}$ target site (TS) bracketing its targeting module (*orange boxes*). Combining Cas9^{D10A}:gRNA$^{R5.2}$ and Cas9^{D10A}:gRNA$^{R5.1}$ complexes gen-

erates a targeted DSB by nicking on opposite DNA strands (in cis paired nicking strategy). PAMs boxed sequences; *magenta arrowheads*, positions of the DSBs and SSBs generated by Cas9 and Cas9^{D10A}, respectively. **b** Schematics of the *CCR5* genotyping assays. DNA products diagnostic for unedited, edited, and mutagenized CCR5 alleles are indicated. RFLA restriction-fragment length analysis; T7EI mismatch-sensing T7 endonucleases I assay; *half arrows*, primers **c** *CCR5* genotyping assays. Genotyping of *CCR5* sequences by RFLA and T7EI assays in HeLa cells transfected with the indicated plasmid sets. RFLA products specific for unedited and HR-edited *CCR5* alleles are identified by *open and closed arrowheads*, respectively; T7EI digestion products diagnostic for genetic changes induced at CCR5 by HR and NHEJ are equally indicated. The genomic DNA analyses were performed at 3 days post-transfection. **d** Comparing genome-editing strategies based on single vs. dual RGNs. HeLa cells were co-transfected with the indicated plasmids and 3 days later RFLA was performed on their genomic DNA. *Open and solid arrowheads* point to unedited and HR-edited *CCR5* sequences, respectively. **e** Comparing *CCR5* mutagenesis in cells exposed to RGNs inducing DSBs vs. SSBs. HeLa cells were co-transfected with the indicated plasmids and T7EI genotyping assays were carried out 3 days later. T7EI products diagnostic for indel footprints left after NHEJ-mediated DSB repair are pinpointed by the *flat arrowhead*

derance of T7EI-digested products, diagnostic for the cumulative build-up of NHEJ and HR events, was detected in cells subjected to DSB-inducing protocols (**Fig. 6c**, *middle panel*). This outcome is consistent with the prevalence of the former over the latter pathway during the repair of DSBs in mammalian cells[4]. Of note, T7EI-digested products corresponding to the in trans paired nicking protocol should mostly represent HR events as nicking exclusively at *CCR5* (single nick) led to the lowest signals in both genotyping assays (**Fig. 6c**). In a follow-up experiment, in addition to the four experimental conditions tested earlier (**Fig. 6c**), we included in cis paired nicking at *CCR5* by transfecting HeLa cells with pS.Donor^{R5} and pCas9^{D10A} mixed with plasmids expressing the gRNA pair gRNA$^{R5.1}$/gRNA$^{R5.2}$. In agreement with the previous data (**Fig. 2c**), in trans paired nicking induced robust accumulation of HR-specific RFLA products. Importantly, cells exposed to DSB-inducing single and dual RGN complexes had a higher proportion of disrupted *CCR5* alleles when compared to those subjected to the SSB-inducing Cas9^{D10A}:gRNA$^{R5.1}$ complex (**Fig. 6e**). These results confirm that in trans paired nicking can achieve programmable nuclease-assisted genome editing without concomitantly introducing a high mutagenic load into target cell populations.

Discussion

In this study, we have demonstrated that in trans paired nicking based on combining RGN "nickases" with RGN-targetable donors can trigger robust and seamless chromosomal insertion of small and large genetic payloads into specific genomic sequences in human cells without the catalytic induction of DSBs. We speculate that the rate-limiting HR steps of single-stranded DNA invasion, donor–acceptor synaptic formation and heteroduplex expansion are, to a great extent, overcome by coordinated presentation of 3′ termini on both interacting partners after in trans paired nicking. These events are shared by recent working models invoking SSBs as recombination-initiating substrates[30]. In addition, recent experiments indicate the in-

volvement of distinct factors underlying canonical and SSB-induced HR pathways. For instance, recombination between donor DNA and a nicked target sequence can proceed through RAD51/BRAC2-independent pathways[30]. In this regard, the versatility of RGNs for inducing nicks at different positions and strands of HR templates, might constitute a valuable experimental system to dissect SSB-dependent HR pathways and, possibly, further improve genome editing based on in trans paired nicking concepts.

Importantly, we also showed that avoiding the use of DSB-inducing nucleases confers a low mutagenic load to this new genome-editing paradigm. Hence, our research complements and joins those of others on devising high-efficiency genome-editing strategies based on RGN "nickases"[31, 32]. In particular, a recent study has demonstrated that fusing cytidine deaminase and uracil DNA glycosylase activities to $Cas9^{D10A}$ results in a large "base editor" capable of inducing C-to-T substitutions within a ~5 nt target window[31]. Another recent study revealed that cleaving and nicking RGNs expose a DNA flap accessible to single-stranded oligodeoxyribonucleotide (ssODN) annealing[32]. On the basis of this information, rationally designed ssODNs and RGN "nickases" were combined and shown to yield homology-directed gene repair in ~10% of treated 293 reporter cells[32]. An intrinsic limitation of these approaches is, however, their unsuitability for effecting extensive genetic changes. Moreover, the fidelity of "base editors" depends on the absence of extra cytidines within the ~ 5 nt "activity window", while that of coupling RGN "nickases" to ssODNs relies on the lack of adventitious mutations created during synthesis and processing of ssODNs *in vitro* and *in cellula*, respectively[33].

The high specificity and accuracy conferred by in trans paired nicking genome editing coupled to its low mutagenic load should be particularly useful in instances in which the precise genetic manipulation of target cell populations is paramount. Examples include the modeling or the repairing of disease traits in stem/progenitor cells and the unbiased genetic screening of cellular phenotypes based on HR-mediated chromosomal insertion of donor DNA libraries[34]. Of note, however, regardless of the DNA targeting specificity and fidelity attained by a particular genome-editing procedure, there is always the risk for uncontrollable random chromosomal insertion of the exogenous DNA. Clearly, these unwanted events can take place in cells that lack or harbor the intended genetic modification.

We have confirmed that nicking RGNs are significantly less mutagenic than their cleaving counterparts at on-target sequences (**Supplementary Fig. 1**). Moreover, experiments done by others have demonstrated that, when compared to cleaving RGNs, nicking RGNs are also significantly less mutagenic at off-target sites[25]. However, regarding the use of "nickases" specifically, one should caution that SSBs can still trigger some mutagenic events if, for instance, after hitting such lesions, an

advancing replication fork collapses resulting in DSB formation[5]. These outcomes will be most problematic at off-target sites. In this regard, sensitive and unbiased assays allowing the genome-wide detection of nick-induced mutagenesis will be instrumental in the future for determining the mutagenic load of gene-editing protocols based on programmable "nickases". Equally related with off-target activities, programmable "nickases" with improved specificities are in demand. Possible candidates include RGN "nickases" built on recently described high-specificity Cas9 scaffolds such as Sp Cas9-HF1[35] and eSpCas9(1.1)[36]. We anticipate that the simple and versatile in trans paired nicking procedure will be compatible with these latest generation tools and, possibly, with other fast-emerging DNA targeting systems.

Concluding, the performance of genome editing depends on its overall efficiency, specificity and fidelity[1]. In this work, we have shown that testing combinatorial interactions between different types of nucleases and foreign DNA structures, can improve these crucial parameters, expanding the options for high-fidelity genetic manipulation of mammalian cells.

Methods

Cells

Human cervix carcinoma HeLa cells (American Type Culture Collection) and its EGFP expressing single cell-derived clone H27[37] were cultured in Dulbecco's modified Eagle's medium (DMEM; ThermoFisher Scientific) containing 5% fetal bovine serum (FBS; ThermoFisher Scientific). Human embryonic kidney (HEK) 293 T cells (American Type Culture Collection) were maintained in DMEM supplemented with 10% FBS. These cells were kept at 37 °C in an humidified-air 10% CO_2 atmosphere. The human embryonic stem cell (ESC) line H1 (ref. 26; WiCell Research Institute) and the induced pluripotent stem cell (iPSC) lines LUMC0044iCtrl44 and LUMC0044iCtrl44.9 were cultured in pluripotent stem cell (PSC) growth medium in the presence of irradiated ICR mouse embryonic fibroblasts (MEFs), and mechanically passaged by a cut and paste method. The PSC growth medium consisted of DMEM/F12 medium with GlutaMax, 20% KnockOut Serum Replacement (KOSR), 10 mM non-essential amino acids (NEAAs), 25 U ml^{-1} penicillin, 25 μg ml^{-1} of streptomycin (all from ThermoFisher Scientific), and 10 ng ml^{-1} of basic fibroblast growth factor (bFGF; Peprotech). PSCs were cultured at 37 °C in an humidified-air 5% CO_2 atmosphere. The cells used in all the experiments were mycoplasma-free. All human materials were collected based on individual written (parental) informed consent after approval by the "Medical Ethics Committee" of the LUMC (reference numbers: P08-087 and P13-080). The experiments involving human materials were done in accordance with the principles outlined in the "Declaration of Helsinki". All animal experiments were approved by the "Animal Experiments Committee" of the LUMC (reference number: 12133) and were performed following the recommendations and guidelines set by the LUMC and the Dutch "Experiments on Animals Act".

Recombinant DNA

The constructs AU26_pCAG.Cas9 and AU28_pCAG.Cas9^{D10A} express Cas9 and Cas9^{D10A}, respectively, from the hybrid CAGGS promoter[38]. The "two-in-one" plasmids AV15_pCAG.Cas9.gRNA^{S1} and AV44_pCAG.Cas9^{D10A}.gRNA^{S1} encode Cas9 and Cas9^{D10A}, respectively, together with the *AAVS1*-targeting gRNAS1. To serve as a negative control, construct AV13_pCas9.gRNA^{NT} expresses Cas9 and the non-targeting gRNA^{NT}. This gRNA is irrelevant in human cells as it addresses Cas9 proteins to the recognition sequence of the *S. cerevisiae* I-SceI homing endonuclease. The annotated maps and full-length nucleotide sequences of AU26_pCAG.Cas9, AU28_pCAG.Cas9^{D10A}, AV15_pCAG.Cas9.gRNA^{S1}, AV44_pCAG.Cas9^{D10A}.gRNA^{S1}, AV13_pCas9.gRNA^{NT}, and AT61_pCas9^{H840A} can be found in the **Supplementary Figs. 17–22**, respectively). Likewise for the *DMD*-targeting donor plasmids AL05_pDonor^{DMD} (Addgene #100284), AL62_pDonor$^{DMD.TS}$ (Addgene #100287), AC62_pDonor$^{DMD.TS.DR}$ (Addgene #100288), and AZ28_pDonor$^{DMD.TS.IR}$ (**Supplementary Figs. 23–26**, respectively), the *AAVS1*-targeting donor constructs AX44_pS.Donor^{S1} (Addgene #100289; **Supplementary Fig. 27**), AX53_pS.Donor$^{S1.TS}$ (Addgene #100290; **Supplementary Fig. 28**) and the *CCR5*-targeting donor plasmids AY42_pS.Donor^{R5} (Addgene #100291; **Supplementary Fig. 29**) and AY10_pS.Donor$^{R5.TS}$ (Addgene #100292; **Supplementary Fig. 30**). The plasmids hCas9 (ref. 39; #41815) and hCas9_D10A[39] (#41816), herein named pCas9 and pCas9^{D10A}, respectively, were obtained from the Addgene repository. The constructs gRNA_Cloning Vector[39] (#41824), gRNA_AAVS1-T2 (ref. 39; #41818), and gRNA_GFP_T2 (ref. 39; #41820), herein called pgRNA^{Empty}, pgRNA^{S1}, and pgRNA^{GFP1}, respectively, were also acquired from Addgene. The plasmid pgRNA^{Empty} expresses no gRNA, whereas pgRNA^{S1} and pgRNA^{GFP1} express gRNAs addressing Cas9 proteins to *AAVS1* and *EGFP* sequences, respectively. The gRNA expressing plasmids AL08_pgRNA^{DMD} (Addgene #100293), AD19_pgRNA$^{S1.2}$, AD13_pgRNA$^{S1.3}$, L06_pgRNA$^{OUT.1}$, AA44_pgRNA$^{OUT.2}$, X32_pgRNA$^{IN.1}$, AA48_pgRNA$^{IN.2}$, AY22_pgRNA$^{R5.1}$ (Addgene #100294), and AY23_pgRNA$^{R5.2}$ (Addgene #100295), were assembled by inserting the annealed oligonucleotides described in the **Supplementary Table 1** into the BveI-digested gRNA acceptor construct S7_pUC.U6.sgRNA.BveI-stuffer[40]. The plasmids AM51_pUC.U6.gRNA^{NT}, herein called pgRNA^{NT} and Z46_pgRNA^{GFP2} encoding an irrelevant gRNA and an *EGFP*-specific gRNA, respectively, have been described before[40]. The AAVS1-targeting donor constructs pSh.AAVS1.eGFP and pAdV.donorS1/T-TS, herein named pDonor.E^{S1} and pDonor.E$^{S1.TS}$, respectively, have been described elsewhere[6, 41]. The additional set of isogenic donor plasmids pDonor^{S1}, pDonor$^{S1.1\times TS}$, and pDonor$^{S1.TS}$ contain the *AAVS1*-targeting module cloned in the pMOLUC vector backbone (Addgene #12514). The pDonor^{S1} plasmid has no gRNA^{S1} target sites, whereas pDonor$^{S1.1\times TS}$ and pDonor$^{S1.TS}$ have one and two gRNA^{S1} target sites, respectively, next to their *AAVS1*-targeting module. The AX35_pDonor.Turq^{S1} and AX28_pDonor.Turq$^{S1.TS}$ have the same composition of pDonor^{S1} and pDonor$^{S1.TS}$ except that they contain a *mTurquoise2* ORF in place of that of *EGFP*. The AT58_pDonor.37^{S1} and AE32_pDonor.37$^{S1.1\times TS}$ share the same EGFP-encoding expression unit present in pDonor^{S1} and pDonor$^{S1.1xTS}$, respectively. However, they differ from pDonor^{S1} and pDonor$^{S1.1xTS}$ in the spacing between their regions of homology (**Supplementary Fig. 5a**, **b**). The final set of *AAVS1*-targeting donor plasmids AV11_pDonor.

EP^{S1} (Addgene #100296) and AV09_ pDonor.$EP^{S1.TS}$ (Addgene #100297) have the same composition of pDonor^{S1} and pDonor$^{S1.TS}$, respectively, except that they encode $Puro^{R}$.T2A.EGFP in place of EGFP (**Supplementary Fig. 31**). The annotated maps and DNA sequences of the constructs generated for this study were assembled with the aid of SnapGene 3.3.4. Where indicated, plasmid pcDNA3.1 (ThermoFisher Scientific) was used as carrier DNA in transfection experiments.

Cell Transfections

One day before transfection, HeLa cells, H27 cells, and 293 T cells were seeded in wells of 24-well plates (Greiner Bio-One; **Supplementary Tables 2–20**). The transfections were initiated by mixing each of the appropriate plasmids together with 1 mg ml^{-1} of polyethyleneimine (PEI, Polysciences) in 50 µl of a 150 mM NaCl solution (Merck). After ~10 s under vigorous vortexing, the transfection mixtures were incubated for 15 min at room temperature, after which they were directly added to the cell cultures (**Supplementary Tables 2–20**). At 3 days post-transfection, the transfection efficiencies were determined by EGFP-directed flow cytometry. Subsequently, the cells were sub-cultured for at least 2 weeks, for the removal of episomal exogenous DNA, after which stable transfection levels were determined by EGFP-directed flow cytometry.

Prior to the transfection of PSCs, the cells were adapted to passaging as single cells by using TrypLE Select (ThermoFisher Scientific) and 10 µM of the Rho kinase inhibitor Fasudil (LC Laboratories). After 2 to 5 single-cell passages, transfections based on Lipofectamine (ThermoFisher Scientific) were initiated using different experimental conditions as detailed in **Supplementary Tables 21** and **22**. In general, single-cell suspensions were generated and incubated for 5–10 min in Opti-MEM (ThermoFisher Scientific) containing the relevant DNA mixtures and a specific Lipofectamine formulation (**Supplementary Tables 21** and **22**). Next, the cell suspensions were seeded on 1 day-old MEF cultures containing PSC growth medium supplemented with 10 µM Fasudil and lacking antibiotics. At 24 h post-transfection the medium was replaced by complete PSC growth medium. After 3–4 days post-transfection, the PSCs were harvested and the transfection efficiencies were determined by EGFP-directed and TRA-1-81-directed flow cytometry. A fraction of the transfected PSCs were seeded and let to divide on MEF cultures for an additional period of 7 to 10 days, after which stable transfection levels were determined by EGFP-directed and TRA-1-81-directed flow cytometry (see below for details).

Adenoviral vectors

The production, purification and titration of adenoviral vector particles $AdV^{\Delta 2}$P.Cas9.F^{50} and $AdV^{\Delta 2}$U6.gRNA^{S1}.F^{50}, herein dubbed AdV.Cas9 and AdV.gRNA^{S1}, respectively, have been specified before[42]. The same methods were applied to generate and characterize AdV.$Cas9^{D10A}$ particles. The genome of AdV.$Cas9^{D10A}$ differs from that of AdV.Cas9 exclusively at codon 10 of the *Cas9* ORF.

T7 endonuclease I-based genotyping assay

Genotyping assays based on the detection of indels by the mismatch-sensing T7 endonuclease I (T7EI) were used for establishing targeted DSB formation activity at *AAVS1*. To this end, 293 T cells were transfected using PEI as indicated in **Supplementary Table 2**. At 3 days post-transfection, cell pellets were collected for subsequent genomic DNA extraction. Genomic DNA was extracted by using the DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions. The target region was amplified by PCR with GoTaq G2 Flexi DNA Polymerase (Promega) following the manufacture's recommendations. The composition of the PCR mixtures and the PCR cycling parameters are specified in **Supplementary Tables 23** and **24**, respectively. Next, the resulting 403 bp amplicons were subjected to T7EI assays[42]. In brief, PCR amplicons were first denatured and reannealed by applying the thermocycler program presented in **Supplementary Table 25**. Afterwards, 10-μl samples were incubated at 37 °C for 17 min in a 15-μl solution containing 1.5 μl of 10 × NEBuffer 2, 0.5 μl of 10 U μl^{-1} T7EI (New England Biolabs) and Milli-Q water. After agarose gel electrophoresis, the DNA fragments were stained with ethidium bromide and were imaged by using a Molecular Imager Gel-Doc™ XR+ apparatus together with the ImageLab 4.1 software (both from Bio-Rad).

Detection of DSBs after AdV-mediate delivery of RGNs

Hela cells were seeded in wells of 24-well plates at a concentration of 6×10^4 cells per well. The next day, they were transduced with different amounts and combinations of adenoviral vectors encoding Cas9, Cas9^{D10A}, or gRNA[51] (**Supplementary Fig. 1**). Mock-transduced cells and cells transduced exclusively with each of these vectors alone served as negative controls. Transduction experiments were carried out in duplicate to generate parallel samples for genotyping and western blot analysis. At 3 days post-transduction, genomic DNA was extracted and T7EI-based genotyping assays were performed as detailed under the previous section. Likewise, at 3 days post-transduction, protein lysates were prepared under ice-cold conditions for western blot analysis as follows. The cells were first washed twice with phosphate-buffered saline (PBS; pH 7.4), after which 60 μl of lysis buffer consisting of 10 mM Tris-HCl (pH 7.5), 1% (v/v) Nonidet P-40, 150 mM NaCl, 0.1% sodium dodecyl sulphate(SDS), and 10 mg ml^{-1} sodium deoxycholate, was added onto the cells for 25 min under gentle plate tilting. The lysis buffer was supplemented with the protease inhibitors present in the cOmplete™, Mini Protease Inhibitor Cocktail Roche-11836153001 (Sigma-Aldrich). Protein concentrations were determined by using the BCA protein assay kit (ThermoFisher Scientific) and a Precisely 1420 Multilabel Plate Counter (PerkinElmer) with the absorption wavelength set at λ = 545 nm. Next, 5-μg protein samples were diluted in bromophenol blue-containing loading buffer consisting of 187.5 mM Tris (pH 6.8), 30% (v/v) glycerol, 9% (w/v) SDS, and 7.5% (v/v) β-mercapthoethanol. Next, the protein samples were heated at 95 °C for 5 min and were resolved through a SDS-8% polyacrylamide gel. After overnight electro-blotting onto Immobilon-P membranes (Millipore), the subsequent blocking, antibody incubation and chemiluminescence protein detection steps, were performed essentially as detailed elsewhere[43] using a monoclonal antibody specific for *S. pyogenes* Cas9 (Diagenode; clone 22B5) and a goat anti-mouse horseradish peroxidase-conjugated IgG (Santa Cruz; sc-2005). These primary and

secondary antibodies were diluted 1:2000 and 1:10,000 in blocking buffer, respectively. To provide for an internal protein loading control, an antibody specific for the α/β tubulin heterodimer (Cell Signalling Technology; 2148) was combined with the aforementioned secondary antibody. These primary and secondary antibodies were diluted 1:5000 and 1:10,000 in blocking buffer, respectively.

Sanger sequencing

The amplicons specific for translocation events between *AAVS1* and *DMD* sequences (**Supplementary Fig. 1e**) and for *AAVS1*-exogenous DNA junctions (**Supplementary Fig. 8**) were amplified, isolated and purified from agarose gel by using the JETquick Gel Extraction Spin kit (Genomed) according to the manufacturer's recommendations. The PCR mixtures and cycling conditions used are presented in **Supplementary Tables 23** and **24**, respectively. Next, the recovered fragments were inserted into pJET1.2/blunt cloning vector provided in the CloneJET PCR Cloning Kit (ThermoFisher Scientific) following the manufacturer's instructions. After transformation, randomly selected clones were grown and subjected to Sanger sequencing (Baseclear, Leiden, the Netherlands). The *AAVS1*-specific PCR products derived from randomly selected iPSC clones (**Fig. 3f**) were purified and subjected to Sanger sequencing for identifying indel footprints generated after RGN activity. All nucleotide sequence reads were aligned and analyzed with the aid of AlignX, Vector NTI Advance R_11.5.0 software. The Sanger sequencing chromatograms were generated by using the Chromas Lite 2.1.1 software (Technelysium Pty).

Flow cytometry

The frequencies of cells expressing EGFP, mTurquoise2 and/or the TRA-1-81 antigen, characteristic of uncommitted PSCs, were determined by using a BD LSR II flow cytometer (BD Biosciences). The TRA-1-81 labeling was carried out by incubating single-cell suspensions of PSCs with a phycoerythrin-conjugated TRA-1-81 antibody (eBioscience) diluted 1:100 in a buffer consisting of PBS supplemented with 0.5% BSA and 2 mM ethylenediaminetetraacetic acid (EDTA). After an incubation period of 30 min at 4 °C in the dark, excess antibody was removed by thorough successive washes with large volumes of the aforementioned buffer. Data was analyzed with the aid FlowJo 7.2.2 software (Tree Star). Non-transfected cells were used to set background fluorescence levels. At least 10,000 events, each representing a single viable cell, were measured per sample.

PCR analyses of gene-editing experiments

The composition of the PCR mixtures and thermocycling parameters used for the analyses of genome-modifying events are discriminated in the **Supplementary Tables 23**, **24**, **26** and **27**. These analyses were performed on whole target cell population as well as on individually sorted cells. The sorting of EGFP+ HeLa and 293 T cells was conducted by using a BD FACSAria III flow cytometer (BD Biosciences) after sub-culturing transfected cell populations for more than 20 days. EGFP+ cells were collected in a 1:1 mixture of regular culture medium containing 2 × penicillin–streptomycin (ThermoFisher Scientific) and FBS. The sorted, EGFP+ cells, were seeded at a density of 0.3 cells per well in wells of 96-well plates (Greiner Bio-

One). To increase the efficiency of cell cloning, the culture media were supplemented with 50 μM α-thioglycerol and 20 nM bathocuprione disulphonate (both from Sigma-Aldrich)[44]. At ~3 weeks after seeding, single cell-derived clones were randomly collected for genomic DNA analysis by junction PCR with the Phire Tissue Direct PCR Master Mix (ThermoFisher Scientific) according to the manufacturer's recommendations. The junction PCR screening of randomly selected puromycin-resistant iPSC colonies was also done by direct PCR. The various direct PCR conditions used in each experiment are specified in **Supplementary Tables 26** and **27**.

Generation of iPSCs

Human fetal fibroblasts were isolated and reprogrammed to iPSCs as detailed elsewhere[45]. In brief, the cell reprogramming was induced by transducing 2×10^4 human fetal fibroblasts seeded in a 12-well plate with the multi-cistronic lentiviral vector LV.RRL.PPT.SF.hOKSM.idTomato.-preFRT[45, 46]. The vector particles, encoding OCT3/4, KLF4, SOX2, and cMYC, were removed 24 h later. At 6 days post-transduction, the cells were harvested and 10^4 of them were seeded on a 10-cm dish with 2×10^6 MEFs cultured in KOSR PSC growth medium (ThermoFisher Scientific). The medium was replenished every other day until the appearance of ESC-like colonies. The resulting iPSCs were subsequently cultured on irradiated ICR MEFs in complete PSC growth medium. The iPSC clone LUMC0044iCtrl44, containing a single provirus, was selected. Subsequently, the chromosomally inserted provirus was removed through hcAd.FLPe.F50-mediated expression of *FLPe* recombinase[43]. To this end, LUMC0044iCtrl44 iPSCs were transduced with hcAd.FLPe.F50 at a multiplicity of infection (MOI) of 20 transduction units/ml. The resulting provirus-free iPSC clone LUMC0044iCtrl44.9 was characterized by immunofluorescence microscopy, COBRA-FISH karyotyping and teratoma assays as described below.

Differentiation of iPSCs

The *in vitro* spontaneous differentiation of iPSCs into the three embryonic germ layers was triggered by culturing clumps of iPSCs on coverslips coated with Vitronectin XF (StemCell Technologies). The cells were incubated in PSC growth medium devoid of bFGF and containing 20% FBS in place of KOSR. The medium was replenished every 2 days. After 3 weeks under differentiation conditions, the cells were stained for the endoderm, mesoderm and ectoderm markers alpha-fetoprotein (AFP), platelet and endothelial cell adhesion molecule 1 (CD31), and tubulin beta 3 class III (TUBB3), respectively. The *in vivo* differentiation of iPSCs into cell types belonging to the three different embryonic germ layers was assessed through teratoma formation assays. To this end, immunodeficient female NOD.Cg-*Prkdc* ^{scid} *Il2rg* ^{tm1Wjl}/SzJ (NSG) mice (8–12 weeks old) were injected subcutaneously with 1×10^6 cells. After 10–16 weeks post-injection, teratomas were isolated and fixed in 4% paraformaldehyde (PFA) and, subsequently, embedded in paraffin and sectioned. Finally, pigmented epithelium (ectoderm), trachea-like epithelium (endoderm), and cartilage (mesoderm) present in iPSC-derived teratomas were identified by hematoxylin–phloxine–saffron (HPS) staining and standard visible light microscopy.

Detection of apoptosis

Two days after transfecting H1 ESCs with Lipofectamine 3000™ (ThermoFisher Scientific) mixed with different construct combinations (**Supplementary Table 22**), the frequencies of apoptotic cells were measured by using Annexin-V Apoptosis Detection Kit-eFluro 450 according to the manufacturer's instructions (ThermoFisher Scientific). H1 ESC cultures treated for 2 h with 1 μg ml^{-1} of staurosporine, stained, or not stained with fluorochrome-conjugated Annexin V, served as positive and negative controls, respectively.

Immunofluorescence microscopy

The acquisition of pluripotency markers (i.e., TRA-1-81, SSEA4, OCT3/4, and NANOG) and differentiation markers (i.e., AFP, CD31, and TUBB3) by iPSCs and iPSC-derived cells, respectively, was assessed via immunofluorescence staining. In brief, cells were first fixed for 15 min in 4% PFA. After several washes with PBS, they were exposed for 1 h to a PBS solution containing 0.1% Triton X-100 and 4% normal swine serum (NSS; Jackson ImmunoResearch). Next, the cells were incubated overnight at 4 °C with primary antibodies directed against TRA-1-81 (1:125; Biolegend Cat. #330702), SSEA4 (1:30; Biolegend Cat. #330402), OCT3/4 (1:100; Santa Cruz Biotechnology Cat. #sc-5279); NANOG (1:500; R&D Systems Cat. #963488); AFP (1:25; Quartett Cat. #2011200530), CD31 (1:100; Dako Cat. #M0823), and TUBB3 (1:4000; Covance Cat. #MMS-435P). These antibodies were diluted in PBS containing 4% NSS. After three 10-min washes with PBS containing 0.05% Tween-20, the specimens were incubated for 1 h in the dark with secondary antibodies (ThermoFisher Scientific) conjugated with Alexa Fluor 488 (1:500) or with Alexa Fluor 568. For the detection of pluripotency and differentiation markers, the Alexa Fluor 568 secondary antibody was applied at 1:200 and 1:500 dilutions, respectively. Nuclei were stained for 5 min in the dark with 4′,6-diamidino-2-phenylindole, dihydrochloride (DAPI; ThermoFisher Scientific). The DAPI was diluted 1:1000 and 1:200 in PBS for the detection of nuclei of pluripotent and differentiated cells, respectively. Finally, the specimens were rinsed with sterile Milli-Q water (Millipore) and were mounted with Mowiol (Calbiochem). The *in vitro* PSC differentiation analyses were performed with an IX51 inverse fluorescence microscope equipped with a XC30 Peltier-cooled digital color camera (Olympus) or a Confocal laser scanning microscope TCS SP8 (Leica). The images were processed with the aid of Cell^{F} 3.4 imaging software (Olympus) or LAS AF software (Leica).

COBRA-FISH karyotyping

Combined binary ratio labeling (COBRA)-FISH analysis was carried out for determining the karyotype of iPSCs essentially following the instructions indicated in a previously published protocol[47]. In short, slides with metaphase chromosomes from iPSCs were pre-treated with RNase I (Roche; 100 μg ml^{-1} in 2 × SSC) and pepsin (Sigma-Aldrich; 0.005% in 0.01 M HCl) at 37 °C for 10 min and 5 min, respectively. After a 10-min fixation with 1% (v/v) formaldehyde in PBS, the specimens were dehydrated by sequential 3-min incubations in 70%, 90% and 100% ethanol. Next, whole-chromosome painting probes, labeled with the fluorescent dyes diethylaminocoumarin, Cy3, Cy5, and rhodamine green using the Universal Linkage System (ULS) kit (Kreatech Biotechnology), were applied to the air-dried slides. After a denaturation

step at 80 °C for 45–90 s, DNA hybridizations were let to proceed in a humidified chamber at 37 °C for 2 days. The unbound probes were removed by a series of post-hybridization washes and the samples were subsequently dehydrated by exposing them to increasing concentrations of ethanol as aforementioned. Finally, the chromosome specimens were sealed and counterstained with Citifluor AF1 mounting solution (Citifluor Ltd.) containing 500 ng ml^{-1} of the DNA dye DAPI. Digital images were acquired with the aid of a Leica DMRA fluorescence microscope coupled to a CCD camera.

Colony-formation assays

In addition to flow cytometric analysis, stable transfections resulting from gene targeting experiments with pDonor.EP^{S1} and pDonor.EP$^{S1.TS}$ were also assessed by colony-formation assays. These assays were applied to puromycin-resistant HeLa and PSC colonies by using, respectively, a standard Giemsa staining and the leukocyte AP kit for detecting alkaline phosphatase (AP) activity (Sigma-Aldrich). These colonies were derived from cell cultures initially exposed to pDonor.EP^{S1} and pCAG.Cas9.gRNA^{S1} (standard setting) or to pDonor.EP$^{S1.TS}$ and pCAG.Cas9^{D10A}.gRNA^{S1} (Nick2 setting). In brief, at 20 days post-transfection, HeLa cells were seeded at densities of 1×10^3 and 1×10^4 cells per 10-cm dish (Greiner Bio-One) in the presence of 1 μg ml^{-1} of puromycin. Puromycin-resistant HeLa cell colonies were identified by Giemsa staining 9 days later. Next to this, at 3–4 days post-transfection PSCs were seeded at a density of 2–3 × 10^5 cells per MEF culture and, 1 day later, were exposed to puromycin at a final concentration of 1 μg ml^{-1}. After 5 days under puromycin selection was assessed by using the aforementioned AP detection kit following the manufacturer's instructions.

Southern blot analysis

Genomic DNA was extracted from individual EGFP+/mTurquoise2+ HeLa cell clones and from a control EGFP+ HeLa cell clone according to a standard organic solvent-based protocol as follows. The cells were collected and incubated overnight at 50 °C in 250 μl of lysis buffer containing 10 mM Tris-HCl (pH 8), 25 mM EDTA (pH 8), 0.5% (w/v) SDS, and 100 mM NaCl. Prior to use, the lysis buffer was supplemented with freshly added proteinase K (ThermoFisher Scientific) at a final concentration of 0.1 μg μl^{-1}. The resulting cell lysates were extracted twice by gentle pipetting in a 1:1 mixture with buffer-saturated phenol:chloroform:isoamyl alcohol (25:24:1). The aqueous phase was recovered and was subsequently subjected to one additional extraction cycle by gentle pipetting in a 1:1 mixture with chloroform. Next, the chromosomal DNA present in the aqueous phase was precipitated in 2.0 and 0.5 volumes of ethanol and 7.5 M ammonium acetate (pH 5.5), respectively. The recovered genomic DNA pellets were washed with 70% (v/v) ethanol, gently air-dried, and were finally dissolved in DNase-free sterile water at a concentration of 1–2 μg μl^{-1}. Subsequently, DNA samples (10 μg each) were digested overnight with BlpI (New England BioLabs) and were resolved through a 1.0% 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 an alkaline transfer buffer consisting of 0.4 N NaOH and 1 M NaCl. After overnight transfer, the membrane was neutralized with a pH 7.2 solution containing 0.5 M Tris-HCl and 1 M NaCl.

The *EGFP*-specific and *mTurquoise2*-specific probes (994 bp each) were isolated from agarose gel after AgeI/HindIII double digestion of AA63_pDonor[S1] and AX28_pDonor.Turq[S1], respectively. The purified DNA probes were radiolabeled with [α-^{32}P]dATP (GE Healthcare Life Sciences) by using the DecaLabel DNA labeling Kit following the manufacturer's instructions (ThermoFisher Scientific). The Pre-hybridization and hybridization steps, 2 h and overnight, respectively, were performed at 65 °C in Rapid-Hyb Buffer (GE Healthcare). Next, the membrane was washed at 65 °C once with a 2 × SSC solution supplemented with 0.1% (w/v) SDS (20 min) and twice with a 0.5 × SSC solution supplemented with 0.1% (w/v) SDS (20 min each). Finally, the membrane was gently air-dried, wrapped in Saran film and exposed to an Amersham Hyperfilm MP (GE Healthcare). The autoradiogram film was obtained by using standard developing solutions.

Statistical analyses

The researchers were not blinded to sample allocation during experiments and data analyses. One-way ANOVA combined with Bonferroni tests were used for the statistical analyses of data sets obtained from three independent experiments comparing the performance of donors with no, one, or two gRNA target sites ($P < 0.05$ was considered significant). The IBM SPSS Statistics 23 software package was employed for these analyses. The comparison of data sets retrieved from standard and in trans paired nicking gene targeting experiments was performed by applying two-tailed Student's *t*-tests ($P < 0.05$ was considered significant). The GraphPad Prism 6 software package was used for these analyses.

Restriction-fragment length analyses

Amplicons spanning the *AAVS1* and *CCR5* target sites, and the BlpI polymorphism in the *mTurquoise2* coding sequence were generated with the PCR reagents and subsequently exposed to the restriction enzymes specified in **Supplementary Table 28**. The corresponding primers and PCR cycling conditions are indicated in **Supplementary Tables 23**, **24**, **26**, and **27**.

Data availability

All relevant results generated in this study are available within the paper and respective Supplementary Information or are available from the corresponding author on reasonable request. Sanger sequencing chromatograms are deposited in FigShare at 10.6084/m9.figshare.5208766.

Acknowledgements

The authors thank Hester Boersma and Daniela Salvatori for their technical assistance with the teratoma assays and Karoly Szuhai and Danielle de Jong for their FISH karyotyping of iPSCs (all from the Leiden University Medical Center, Leiden, the Netherlands). Axel Schambach (Institute of Experimental Hematology, Hannover, Germany), Anna Gram (Leiden University Medical Center, Leiden, the Netherlands), and Alessandra Recchia (University of Modena and Reggio Emilia, Modena, Italy) are gratefully acknowledged for providing the construct for generating iPSCs, the PuroR.T2A.EGFP ORF and the pSh.AAVS1.eGFP donor plasmid (herein named pDonor.E[S1]), respectively. The authors also thank Rob Hoeben for his critical reading of the manuscript and Steve Cramer for his assistance with the logistics

(both from the Leiden University Medical Center, Leiden, the Netherlands). The authors thank Ron Wolterbeek (Department of Medical Statistics and Bioinformatics, Leiden University Medical Center) for his help with the statistical analysis using the IBM SPSS Statistics 23 software package. This work was supported by the Dutch Prinses Beatrix Spierfonds (W.OR11-18) and the French AFMTéléthon (16621). X.C. is the recipient of a Ph.D. research fellowship from the China Scholarship Council-Leiden University Joint Scholarship Programme.

Supplementary information

Supplementary information an be found at doi:10.1038/s41467-017-00687-1.

Competing interests

The authors declare no competing financial interests.

References

1. Maggio I, Gonçalves MA. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 2015;33:280–291. doi: 10.1016/j.tibtech.2015.02.011.
2. Bolukbasi MF, Gupta A, Wolfe SA. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. Nat. Methods. 2016;13:41–50. doi: 10.1038/nmeth.3684.
3. Wright AV, Nunez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell. 2016;164:29–44. doi: 10.1016/j.cell.2015.12.035.
4. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. Nat. Rev. Mol. Cell Biol. 2003;4:712–720. doi: 10.1038/nrm1202.
5. Cannan WJ, Pederson DS. Mechanisms and consequences of double-strand DNA break formation in chromatin. J. Cell. Physiol. 2016;231:3–14. doi: 10.1002/jcp.25048.
6. Holkers M, et al. Adenoviral vector DNA for accurate genome editing with engineered nucleases. Nat. Methods. 2014;11:1051–1057. doi: 10.1038/nmeth.3075.
7. Lackner DH, et al. A generic strategy for CRISPR-Cas9-mediated gene tagging. Nat. Commun. 2015;6:10237. doi: 10.1038/ncomms10237.
8. He X, et al. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res. 2016;44:e85. doi: 10.1093/nar/gkw064.
9. Suzuki K, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature. 2016;540:144–149. doi: 10.1038/nature20565.
10. Chu VT, Weber T, Wefers B. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. 2015;33:543–548. doi: 10.1038/nbt.3198.
11. Maruyama T, et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. 2015;33:538–542. doi: 10.1038/nbt.3190.
12. Robert F, Barbeau M, Ethier S, Dostie J, Pelletier J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. 2015;7:93. doi: 10.1186/s13073-015-0215-6.
13. 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. 2009;106:5099–5104. doi: 10.1073/pnas.0810588106.
14. van Nierop GP, de Vries AA, Holkers M, Vrijsen KR, Gonçalves MA. Stimulation of ho-

mology-directed gene targeting at an endogenous human locus by a nicking endonuclease. Nucleic Acids Res. 2009;37:5725–5736. doi: 10.1093/nar/gkp643.
15. Metzger MJ, McConnell-Smith A, Stoddard BL, Miller AD. Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. Nucleic Acids Res. 2011;39:926–935. doi: 10.1093/nar/gkq826.
16. Ramirez CL, et al. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucleic Acids Res. 2012;40:5560–5568. doi: 10.1093/nar/gks179.
17. Wang J, et al. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. Genome Res. 2012;22:1316–1326. doi: 10.1101/gr.122879.111.
18. Papapetrou EP, Schambach A. Gene insertion into genomic safe harbors for human gene therapy. Mol. Ther. 2016;24:678–684. doi: 10.1038/mt.2016.38.
19. Mendell JR, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann. Neurol. 2012;71:304–313. doi: 10.1002/ana.23528.
20. Maggio I, Chen X, Gonçalves MA. The emerging role of viral vectors as vehicles for DMD gene editing. Genome Med. 2016;8:59. doi: 10.1186/s13073-016-0316-x.
21. Hockemeyer D, et al. Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. 2011;29:731–734. doi: 10.1038/nbt.1927.
22. Holliday R. A mechanism for gene conversion in fungi. Genet. Res. 1964;5:282–304. doi: 10.1017/S0016672300001233.
23. Gonçalves MA, van Nierop GP, Holkers M, de Vries AA. Concerted nicking of donor and chromosomal acceptor DNA promotes homology-directed gene targeting in human cells. Nucleic Acids Res. 2012;40:3443–3455. doi: 10.1093/nar/gkr1234.
24. Mali P, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 2013;31:833–838. doi: 10.1038/nbt.2675.
25. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154:1380–1389. doi: 10.1016/j.cell.2013.08.021.
26. Thomson JA, et al. Embryonic stem cell lines derived from human blastocysts. Science. 1998;282:1145–1147. doi: 10.1126/science.282.5391.1145.
27. Liu JC, Lerou PH, Lahav G. Stem cells: balancing resistance and sensitivity to DNA damage. Trends Cell. Biol. 2014;24:268–274. doi: 10.1016/j.tcb.2014.03.002.
28. Merkle FT, et al. Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. Cell Rep. 2015;11:875–883. doi: 10.1016/j.celrep.2015.04.007.
29. Wang CX, Cannon PM. The clinical applications of genome editing in HIV. Blood. 2016;127:2546–2552. doi: 10.1182/blood-2016-01-678144.
30. Davis L, Maizels N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. Proc. Natl Acad. Sci. USA. 2014;111:E924–E932. doi: 10.1073/pnas.1400236111.
31. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533:420–424. doi: 10.1038/nature17946.
32. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat. Biotechnol. 2016;34:339–344. doi: 10.1038/nbt.3481.
33. Chen F, et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat. Methods. 2011;8:753–755. doi: 10.1038/nmeth.1653.
34. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. Nature. 2014;513:120–123. doi: 10.1038/

nature13695.
35. Kleinstiver BP, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529:490–495. doi: 10.1038/nature16526.
36. Slaymaker IM, et al. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351:84–88. doi: 10.1126/science.aad5227.
37. Gonçalves MA, van der Velde I, Knaan-Shanzer S, Valerio D, de Vries AA. Stable transduction of large DNA by high-capacity adeno-associated virus/adenovirus hybrid vectors. Virology. 2004;321:287–296. doi: 10.1016/j.virol.2004.01.007.
38. Niwa H, Yamamura K, Miyazaki J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene. 1991;108:193–199. doi: 10.1016/0378-1119(91)90434-D.
39. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–826. doi: 10.1126/science.1232033.
40. Chen X, et al. Probing the impact of chromatin conformation on genome editing tools. Nucleic Acids Res. 2016;44:6482–6492. doi: 10.1093/nar/gkw524.
41. Coluccio A, et al. Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. Mol. Ther. 2013;21:1695–1704. doi: 10.1038/mt.2013.143.
42. Maggio I, et al. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Sci. Rep. 2014;4:5105. doi: 10.1038/srep05105.
43. Gonçalves MA, et al. Targeted chromosomal insertion of large DNA into the human genome by a fiber-modified high-capacity adenovirus-based vector system. PLoS ONE. 2008;3:e3084. doi: 10.1371/journal.pone.0003084.
44. Brielmeier M, et al. Improving stable transfection efficiency: antioxidants dramatically improve the outgrowth of clones under dominant marker selection. Nucleic Acids Res. 1998;26:2082–2085. doi: 10.1093/nar/26.9.2082.
45. Dambrot C, et al. Polycistronic lentivirus induced pluripotent stem cells from skin biopsies after long term storage, blood outgrowth endothelial cells and cells from milk teeth. Differentiation. 2013;85:101–109. doi: 10.1016/j.diff.2013.01.001.
46. Warlich E, et al. Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. Mol. Ther. 2011;19:782–789. doi: 10.1038/mt.2010.314.
47. Szuhai K, Tanke HJ. COBRA: combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. Nat. Protoc. 2006;1:264–275. doi: 10.1038/nprot.2006.41.

Chapter 3

Probing the **Impact** of **Chromatin** Conformation on **Genome Editing** Tools.

Nucleic Acids Res. 2016 Jul 27;44(13):6482-92.
Chen X, Rinsma M, Janssen JM, Liu J, Maggio I, Gonçalves MA.

Abstract

Transcription activator-like effector nucleases (TALENs) and RNA-guided nucleases derived from clustered, regularly interspaced, short palindromic repeats (CRISPR)-Cas9 systems have become ubiquitous genome editing tools. Despite this, the impact that distinct high-order chromatin conformations have on these sequence-specific designer nucleases is, presently, ill-defined. The same applies to the relative performance of TALENs and CRISPR/Cas9 nucleases at isogenic target sequences subjected to different epigenetic modifications. Here, to address these gaps in our knowledge, we have implemented quantitative cellular systems based on genetic reporters in which the euchromatic and heterochromatic statuses of designer nuclease target sites are stringently controlled by small-molecule drug availability. By using these systems, we demonstrate that TALENs and CRISPR/Cas9 nucleases are both significantly affected by the high-order epigenetic context of their target sequences. In addition, this outcome could also be ascertained for *S. pyogenes* CRISPR/Cas9 complexes harbouring Cas9 variants whose DNA cleaving specificities are superior to that of the wild-type Cas9 protein. Thus, the herein investigated cellular models will serve as valuable functional readouts for screening and assessing the role of chromatin on designer nucleases based on different platforms or with different architectures or compositions.

Introduction

Transcription activator-like effector (TALE) nucleases (TALENs) and RNA-guided clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated Cas9 (CRISPR/Cas9) nucleases, have become the prevalent tools for inducing targeted double-stranded DNA breaks (DSBs) in living cells [1,2].

Native TALE proteins have evolved in phytopathogenic bacteria (*Xanthomonas sp.*) to serve as transcriptional activators of specific host plant genes whose products promote infection [1,2]. Typically, TALE DNA-binding domains consist of an array of 15.5–19.5 repeats composed of 33–34 residues containing polymorphisms called repeat variable di-residues (RVDs) at positions 12 and 13. Individual RVDs mediate the binding of the repeat in which they are embedded to a particular nucleotide. This direct one-to-one interaction between nucleotides and repeats permits the straightforward assembling of artificial proteins, among which TALENs, with specific DNA binding activities. Indeed, TALENs are built by fusing the DNA-binding programmable polymorphic repeats from TALE proteins to the nuclease domain of the type IIS restriction enzyme FokI [1,2]. The recognition of preselected genomic sequences by TALEN pairs leads to *in situ* FokI dimerization resulting in nuclease activation and targeted DSB formation.

Native CRISPR/Cas9 nucleases have evolved in bacteria and archaea as adaptive immune systems to fend off invading nucleic acids (e.g. bacteriophage and plasmid DNA) whose chromatin signatures are, clearly, fundamentally different from those present in and acquired by eukaryotic nuclear genes [3]. Programmable CRISPR/Cas9 nucleases are ribonucleoprotein complexes composed of a sequence-tailored single guide RNA (gRNA) and a Cas9 protein harboring two nuclease domains (i.e. RuvC and HNH). The 5′ and 3′ ends of the gRNAs serve as targeting and scaffolding moieties for Cas9, respectively. The initial interaction between CRISPR/Cas9 complexes and DNA involves binding of Cas9 to a nucleotide sequence named protospacer adjacent motif (PAM; NGG in the case of the prototypic Cas9 from *S. pyogenes*). This initial engagement with DNA is mediated through a composite PAM-interacting domain located in two regions of the Cas9 protein [4]. Subsequently, targeted DSB formation is triggered after the hybridization of the 5′ end of the gRNA to a complementary ~20-bp genomic target sequence located next to the PAM [1,2].

Therefore, TALENs and CRISPR/Cas9 nucleases operate in strikingly different manners in that the target site specificity of the former is governed via protein–DNA interactions, whilst that of the latter is ultimately dictated by RNA–DNA hybridizations [1,2]. The activation of cellular DNA repair pathways resulting from the activity of these sequence-specific designer nucleases is being harnessed in an ever-increasing number of genome editing settings. For instance, the repair of targeted DSBs via non-homologous end joining (NHEJ) can result in the incorporation of small insertions and deletions (indels) leading to gene knock-out or gene correction [1,2].

Well-defined parameters that can affect targeted DSB formation by TALENs and CRISPR/Cas9 nucleases *in cellula* include their specific construction, composition, primary target sequence and, in the case of TALENs, CpG methylation [5,6]. In contrast, there has been no direct and quantitative assessment of the impact that high-order chromatin conformations have on these gene-editing tools at isogenic target sites [7]. By the same token, an investigation of the relative performance of TALENs and CRISPR/Cas9 nucleases at target sequences subjected to different epigenetic modifications is equally lacking. Indeed, hitherto, studies based mostly on catalytically 'dead' Cas9 enzymes have exclusively correlated preferential interactions of CRISPR/Cas9 complexes with open chromatin regions bearing candidate off-target sites (e.g. 5-nucleotide seed sequences followed by the *S. pyogenes*'s PAM) [7-11]. In this regard, it is also of note that binding of CRISPR/Cas9 complexes to DNA is, for the most part, uncoupled from actual phosphodiester bond cleavage [7,8,11]. Here, to cover these gaps in our knowledge, we have adapted and validated cellular systems for tracking and measuring the impact of the epigenetically regulated three-dimensional chromatin structure on gene editing processes. By using these reporter systems, we demonstrate that TALENs and CRISPR/Cas9 nucleases are both significantly hindered by the chromatin context in which their target sequences are embedded.

Results and Discussion

To investigate and compare the impact that the chromatin structure has on the performance of TALENs and CRISPR-Cas9 complexes, we have set-up two complementary quantitative cellular systems dubbed HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$. In these systems, the chromatin conformations (i.e. euchromatic versus heterochromatic) at isogenic target sequences are stringently controlled through small-molecule drug availability. In particular, these experimental models based on human cells, harbor reporter alleles whose transition from compact to relaxed chromatin is governed through doxycycline (Dox)-dependent release of a dominantly silencing tTR-KRAB fusion protein from its cognate *TetO* recognition elements (**Fig. 1A**). The Krüppel-associated box domain (KRAB) is the effector moiety of the largest class of transcriptional repressors in vertebrates. The binding of KRAB-containing proteins to DNA triggers epigenetic silencing mechanisms involving the recruitment of, amongst others, scaffolding and chromatin remodeling factors such as KRAB-associated protein 1 (KAP-1) and heterochromatin protein 1 (HP-1) [27,28]. The impinged facultative heterochromatin is characterized by epigenetic marks of silenced genes [29,30].

The first system, HER.TLR$^{TetO.KRAB}$, consists of tTR-KRAB-expressing human embryonic retinoblasts containing a Traffic Light Reporter (*TLR*) construct [12], which we have herein modified by flanking it with *TetO* (*TLR^{TetO}*) elements (**Fig. 1B**). The original *TLR* construct has an out-of-frame *mCherry* reporter located downstream of a *T2A* sequence and an *EGFP* ORF disrupted by an I-SceI recognition site [12]. The repair by NHEJ of DSBs made at sequences upstream of *mCherry* generates indels that can place the nucleotide sequence of *mCherry* in-frame. The resulting red fluorescent cells thus report sequence-specific nuclease activity. To validate the cellular model based on the adapted *TLR* construct *TLR^{TetO}*, ChIP-qPCR assays were performed on HER.TLR$^{TetO.KRAB}$ cells incubated in the presence or in the absence of Dox. In these assays, antibodies directed against histone 3 acetylation (H3Ac) and histone 3 lysin 9 trimethylation (H3K9me3), characteristic of open and closed chromatin, respectively, were used. The ChIP-qPCR analysis established a Dox-dependent switch of *TLR^{TetO}* sequences from a heterochromatic to a euchromatic state (**Fig. 1C**). Indeed, in HER.TLR$^{TetO.KRAB}$ cells not exposed to Dox, the enrichment factors for the heterochromatin mark H3K9me3 at six randomly-selected ChIP-qPCR target regions spanning the *TLR^{TetO}* gene body varied from a minimum of 2.2-fold to a maximum of 14.4-fold when compared to those measured in their Dox-treated counterparts (**Fig. 1C** *left panel*). Conversely, in HER.TLR$^{TetO.KRAB}$ cells exposed to Dox, the enrichment factors for the euchromatin mark H3Ac at the same ChIP-qPCR target regions ranged from a minimum of 1.1-fold to a maximum of 61.7-fold when compared to those measured in their Dox-negative counterparts (**Fig. 1C**, *right panel*). These variations in histone modifications are well within those reported to induce biologi-

cally relevant changes in cells [30,31]. Taken together, these data validate the *TLR*^{TetO} construction as a Dox-dependent epigenetically controlled system that can be used for studying the effect of chromatin conformation on the performance of gene editing tools. The second system, HEK.EGFP$^{TetO.KRAB}$, entails *tTR-KRAB*-expressing human embryonic kidney cells with a *TetO*-flanked functional *EGFP* allele [12] (**Fig. 1D**).

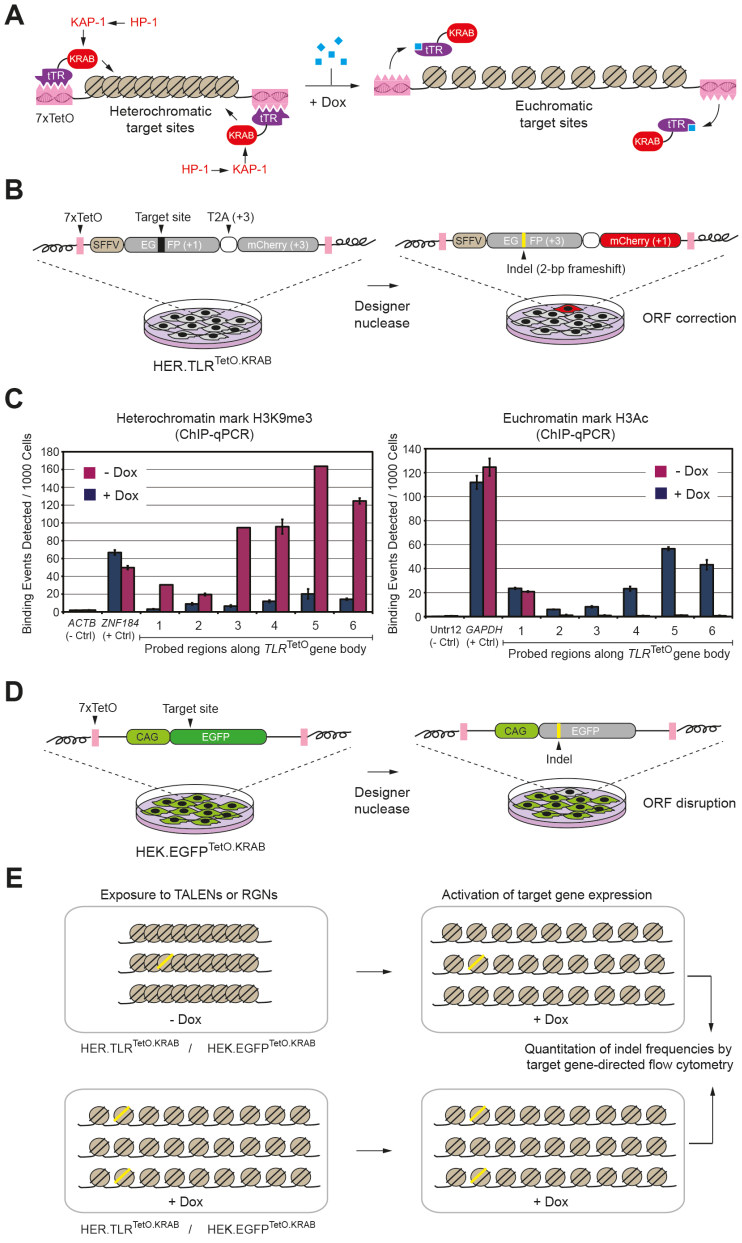


Fig. 1
Experimental systems for tracking designer nuclease-induced indel formation at target sites

with alternate epigenetic states. (**A**) Drug-dependent control over the chromatin conformation of designer nuclease target sites. In this system, the binding of the *trans*-acting tTR-KRAB fusion protein to *cis*-acting *TetO* sequences leads to the recruitment of epigenetics modulators consisting of, amongst others, KAP1 and HP1 proteins. In the presence of doxycycline, tTR-KRAB cannot bind its cognate *TetO* elements, resulting in the transition of a compact heterochromatic to a relaxed euchromatic conformation. (**B**) Designer nuclease-induced gain-of-function system (ORF correction). HER.TLR$^{TetO.KRAB}$ reporter cells contain a *TetO*-flanked *TLR* allele. Subjecting tTR-KRAB-expressing HER.TLR$^{TetO.KRAB}$ cells to designer nucleases targeting *TLR* sequences yields ORF-correcting indels generated by NHEJ-mediated DSB repair and the appearance of mCherry-positive cells. (**C**) ChIP-qPCR analysis on HER.TLR$^{TetO.KRAB}$ cells. ChIP-qPCR signals detected by using antibodies directed against open and closed chromatin marks (H3Ac and H3K9me3, respectively). Six different regions spanning the *TLR^{TetO}* gene body were probed. The targeted sequences were located in the *EF1α* promoter (EF1α), the *puromycin resistance* ORF (Puro^{R}), the spleen focus-forming virus regulatory elements (*SFFV*), the *EGFP* ORF (EGFP a and EGFP b) and the *mCherry* ORF (mCherry). Standard positive and negative controls (Ctrl) are indicated. (**D**) Designer nuclease-induced loss-of-function system (ORF disruption). HEK.EGFP$^{TetO.KRAB}$ reporter cells harbor a *TetO*-flanked *EGFP* target allele. Exposing tTR-KRAB-expressing HEK.EGFPT$^{etO.KRAB}$ cells to designer nucleases targeting *EGFP* yields ORF-disrupting indels generated by NHEJ DSB repair and the emergence of EGFP-negative cells. (**E**) Experimental settings. HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$ cells exposed or not to Dox are transfected with designer nuclease-encoding constructs. After the generation of site-specific DSBs and ensuing NHEJ-mediated indel formation in each of the two parallel settings (*yellow boxes*), target gene expression is activated allowing to quantify the frequencies of NHEJ-based gene editing by flow cytometry.

In this system, the traceable cellular phenotype consists of non-fluorescent cells emerging from the incorporation of frame-shifting indels at *EGFP* sequences after NHEJ-mediated DSB repair. In conclusion, the generation of DSBs within isogenic euchromatic and heterochromatic target sites located in *TLR* and *EGFP* reporter genes can readily be tracked *in cellula* by quantifying mCherry-positive and EGFP-negative cells, respectively (**Fig. 1E**).

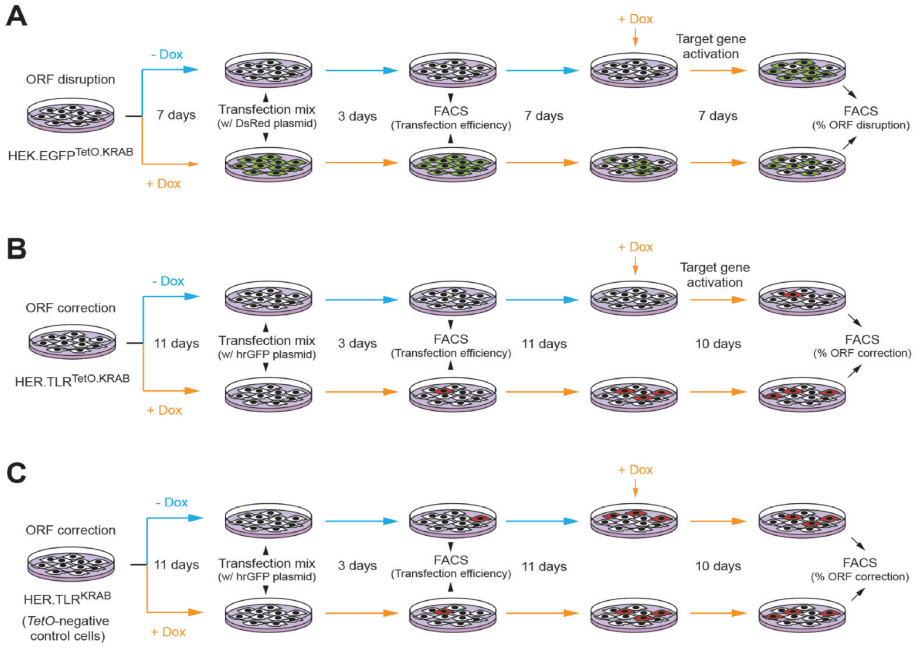


Fig. 2
Detailed diagrammatic representation of the experimental designs used in the present study. The tTR-KRAB-expressing reporter cells HEK.EGFP$^{TetO.KRAB}$(**A**) and HER.TLR$^{TetO.KRAB}$ (**B**) were used for tracking and quantifying designer nuclease-induced gene editing events at target sites subjected to different epigenetic states. The *TetO*-negative and tTR-KRAB-expressing reporter cells HER.TLR^{KRAB}(**C**) were also generated to provide for negative controls. The HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$ systems are complementary in that they allow for measuring ORF disruption and ORF correction, respectively. The initial high-order chromatin conformation of both model alleles is controlled through Dox-dependent regulation of tTR-KRAB binding. Reporter cells HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$, containing target sequences in a heterochromatic (−Dox) or euchromatic (+Dox) state, are transiently transfected with different sets of designer nuclease-encoding constructs. DsRed and hrGFP expression plasmids are included in the transfection mixtures to serve as internal controls for transfection efficiency. After the generation of targeted DSBs in each of the two parallel settings (i.e. −Dox and +Dox), target gene expression is activated allowing to quantifying the frequencies of NHEJ-based gene editing by flow cytometry.

We started by transfecting HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cells, cultured in the presence or in the absence of Dox, with expression plasmids encoding TALENs and CRISPR/Cas9 nucleases (**Supplementary Fig. S5**). Parallel cell cultures transfected with constructs expressing a single TALEN monomer, Cas9 plus a non-targeting gRNA or Cas9 plus an 'empty' gRNA, served as negative controls (Ctrl). After the action of the various designer nucleases had taken place, all cell cultures were exposed to Dox in order to allow for transgene activation and ensuing flow cytometric analysis (**Fig. 1E** and **2**). The resulting data revealed that in HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cells (i.e. ORF correction and ORF disruption readouts, respectively), the frequencies of DSB formation were significantly lower in cells whose target sites were embedded in heterochromatin (-Dox) when compared to those attained when the same sites were located in euchromatin (+Dox) instead (**Fig. 3A–D** and **Supplementary Fig. S6**). Indeed, euchromatic target sequences in reporter HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cells were cleaved by CRISPR/Cas9 complexes on average 2.7- and 2.2-fold, more frequently than when they were in a heterochromatic state (**Fig. 3A** and **B**, respectively). Similarly, euchromatic target sites in reporter HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cells were cut by TALEN dimers on average 5.2- and 3.0-fold more often than their isogenic heterochromatic counterparts (**Fig. 3C** and **D**, respectively). Thus, these ratios between the levels of targeted DSB formation in epigenetically open versus closed DNA can be referred to as the chromatin impact index of a particular nuclease.

Of note, the TALEN pairs TALEN-43-L/TALEN-43-R and TALEN-GA-L/TALEN-GA-R (**Fig. 3C** and **D**, respectively) are based on different architectures consisting of TALE scaffolds from *Xanthomonas axonopodis pv. citri* and *Xanthomonas campestris pv. armoraciae*, respectively [32,33]. Importantly, no differences in designer nuclease-induced targeted DSB frequencies were measured in HER.TLR^{KRAB} cells whose *TLR* sequences are not subjected to conditional KRAB-mediated epigenetic regulation (**Fig. 3E**). HER.TLR^{KRAB} and HER.TLR$^{TetO.KRAB}$ cells differ from each other in that the former

lacks the *cis*-acting *TetO* elements (**Fig. 1A** and **3E**, respectively). In particular, expression of TALEN dimers and CRISPR/Cas9 complexes directed to *TLR* sequences in HER.TLR^{KRAB} cells led to similar levels of site-specific DSB formation in Dox-containing and Dox-free settings resulting in chromatin impact indexes of roughly 1 for both types of designer nucleases (**Fig. 3F** and **G**).

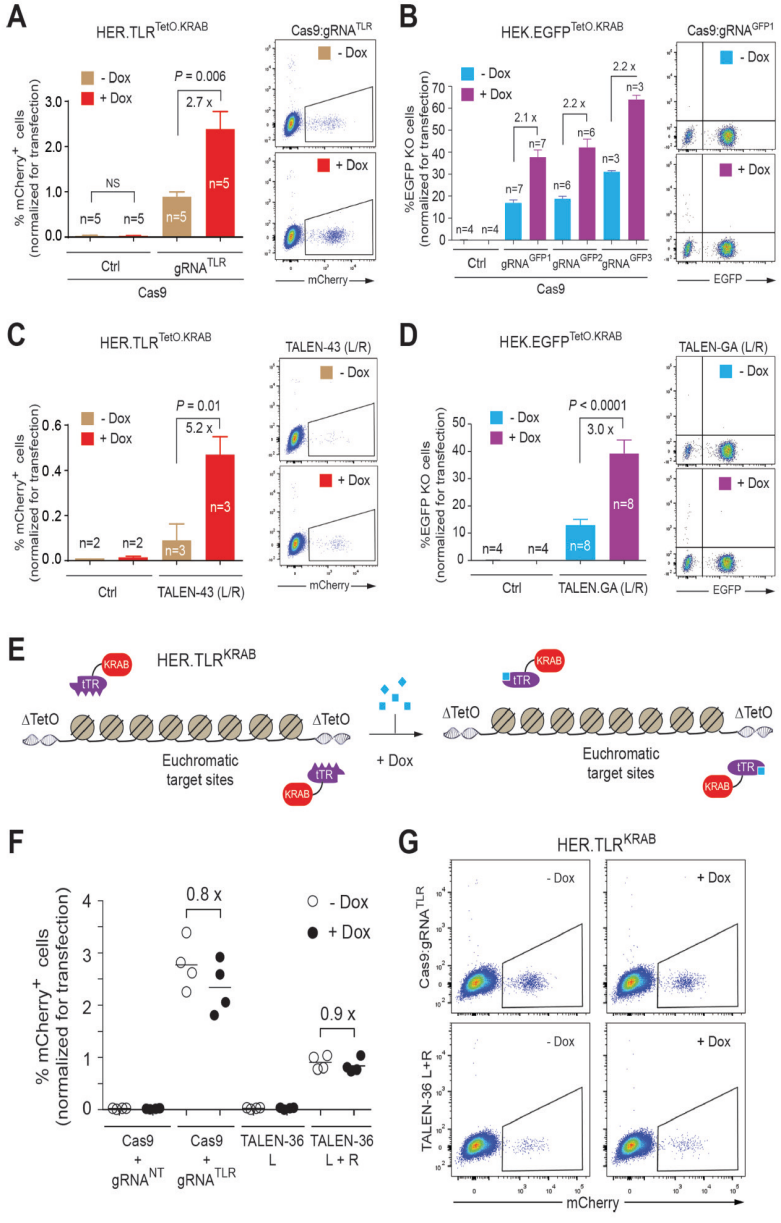


Fig. 3
The impact of distinct chromatin conformations on the frequencies of NHEJ-based gene editing achieved by TALEN and CRISPR/Cas9 nucleases. Reporter cells HER.TLR$^{TetO.KRAB}$ (**A**

and **C**) and HEK.EGFP$^{TetO.KRAB}$ (**B** and **D**), were subjected to the indicated experimental conditions. The negative controls (Ctrl) in panels **A** and **B**, involved transfecting cells with expression plasmids encoding Cas9 mixed with a non-targeting gRNA or with an 'empty' gRNA construct, respectively. The negative controls (Ctrl) in panel **C** and in panel **D** involved exposing target cells exclusively to a single TALEN monomer. Representative flow cytometry dot plots are also presented next to each graph. Ten thousand events, each corresponding to a single viable cell, were measured per sample. Error bars indicated mean ± s.e.m. *P* values (by two-tailed *t*-tests) and the number of independent experiments (*n*) are shown. (**E**) Control HER.TLR^{KRAB} cells. The chromatin status of *TLR* sequences in HER.TLR^{KRAB} cells are not controlled by Dox since they lack *cis*-acting *TetO* elements for tTR-KRAB binding. (**F**) Gene editing in HER.TLR^{KRAB} cells. HER.TLR^{KRAB} cells were either exposed or not to Dox and were subsequently transfected with the indicated constructs. Differences between +Dox and −Dox values were not statistically significant as determined by three-way ANOVA (P = 0.151; two independent experiments done in replicate). The ratios between EGFP knockout levels measured in the presence versus those determined in the absence of Dox. gRNA^{NT}, Non-targeting gRNA. (**G**) Representative flow cytometry dot plots corresponding to the experimental settings presented in panel **F**.

Next, we carried out another series of gene knockout experiments in HEK.EGFP-$^{TetO.KRAB}$ cells by using a panel of TALENs composed of different numbers of TALE repeats in their DNA-binding arrays. Once again, the frequencies of targeted gene knockout were highest in cells containing euchromatic target sites (**Fig. 4A** and **B**). Interestingly, the TALEN complex built on monomers with the shortest TALE DNA-binding domains was the most hindered by having its target sequence embedded in heterochromatin. In fact, under this condition, this TALEN pair led to gene editing levels barely above background (**Fig. 4A**). These data suggest that a minimum TALEN DNA-binding energy is necessary in order to overcome the compact heterochromatic barrier and induce meaningful levels of site-specific DSB formation. Nonetheless, it is important to point out that the majority of TALEN proteins currently in use contain DNA-binding domains made up of well over 14.5 TALE repeats (i.e. typically, 17.5 to 18.5 TALE repeats per monomer) [1,2].

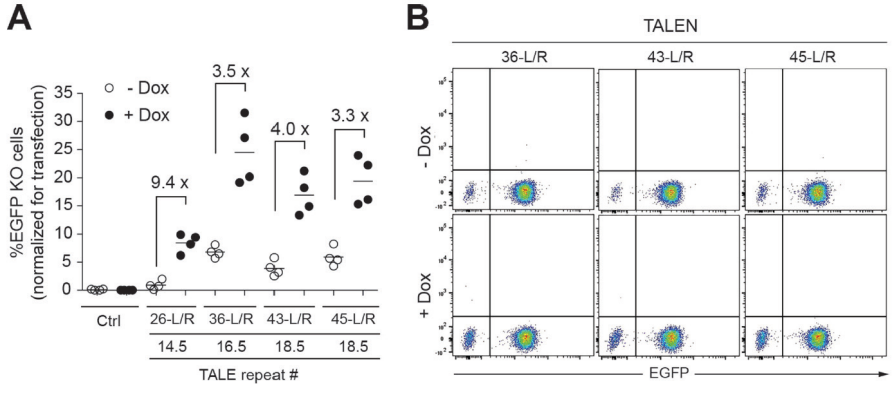


Fig. 4
Gene editing experiments at *EGFP* target sequences subjected to alternative chromatin conformations. (**A**) Targeted mutagenesis induced by TALENs with different numbers of TALE repeats. HEK.EGFP^{TetOKRAB} cells were either incubated or not with Dox and were subsequently

transfected with expression plasmids encoding TALENs with 14.5, 16.5 and 18.5 TALE repeats. Negative controls consisted of parallel cultures exposed exclusively to TALEN-GA-L (Ctrl). After the generation of site-specific DSBs in each of the two parallel settings (i.e. -Dox and +Dox), target gene expression was activated in all cultures by adding Dox for determining the frequencies of NHEJ-mediated *EGFP* knockout by flow cytometry. The ratios between *EGFP* knockout levels measured in the presence versus those determined in the absence of Dox (open versus closed chromatin, respectively) are indicated. (**B**) Representative flow cytometry dot plots corresponding to the experimental settings presented in panel **A**.

Informed by structural data [4,34], two independent research groups have recently generated *S. pyogenes* Cas9 variants with targeting DNA specificities markedly superior to that of their parental wild-type counterpart [19,21]. These high-specificity variants, dubbed SpCas9-HF1 [19] and eSpCas9(1.1) [21], have amino acid substitutions which reduce nonspecific protein–DNA interactions and, as a result, constrain the respective ribonucleoprotein complexes to preferentially cut at the intended target sites. The SpCas9-HF1 mutations N497A, R661A, Q695A and Q926A reduce protein binding to the sugar-phosphate backbone of the complementary strand by removing hydrogen bridges; the charged-to-alanine eSpCas9(1.1) mutations K848A, K1003A and R1060A affect protein binding to the non-complementary stand by eliminating cationic amino acids along the nt-groove located between the HNH and RuvC-like nuclease domains. Regardless of their different point mutations and modus operandi, these engineered Cas9 variants contribute to addressing in a very direct manner a major issue in the genome-editing field, that is, the need for reducing off-target DNA cleaving activities as these can confound experimental outcomes and obstruct potential clinical applications. To determine the relative chromatin impact indexes of these new Cas9 variants, we carried out targeted gene knockout experiments in HEK.EGFP$^{TetO.KRAB}$ cells. In these experiments, constructs encoding wild-type Cas9, SpCas9-HF1 or eSpCas9(1.1) were transfected together with plasmids expressing five different *EGFP*-targeting gRNAs. Data presented in **Fig. 5A** show that, as previously established for wild-type Cas9, high-specificity Cas9 variants were both hindered by heterochromatin. Interestingly, of the two variants, the SpCas9-HF1 protein was the most affected in these experiments as indicated by its higher chromatin impact index (**Fig. 5B**).

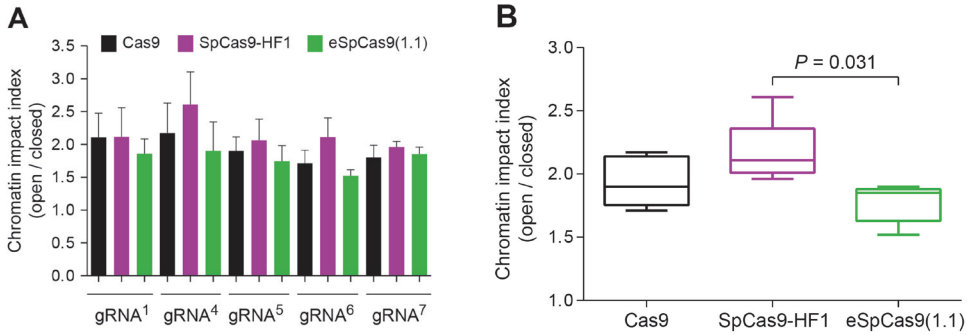


Fig. 5
Testing the impact of chromatin conformation on high-specificity Cas9 variants. (**A**) Screening of CRISPR/Cas9 complexes with different Cas9 proteins in HEK.EGFP^{TetOKRAB} cells. HEK.EGFP^{TetOKRAB} cells were incubated in the presence or in the absence of Dox and were subsequently transfected with expression plasmids encoding the indicated CRISPR/Cas9 nuclease components. The chromatin impact index was determined by computing the ratios between *EGFP* knockout levels measured in the presence versus those determined in the absence of Dox. Error bars indicate mean ± s.e.m. corresponding to three independent experiments. Ten thousand events, each corresponding to a single viable cell, were measured per sample. **(B**) Cumulative chromatin impact indexes. Boxplot of the chromatin impact indexes presented in panel **A**. *Whiskers*, minimum and maximum. One-way ANOVA compared the experimental groups with a subsequent comparison between groups being done by Bonferroni analysis ($P < 0.05$ was considered significant).

Taken together, our data establishes that distinct high-order epigenetically regulated chromatin conformations can have a significant impact on DNA cleaving activity regardless of the type of the designer nuclease tested and of the specific nucleotide sequence targeted. Moreover, of the two main designer nuclease platforms currently in use, TALENs seem to be the most influenced by the high-order chromatin status of their target DNA as revealed by the cumulative data presented in **Fig. 6**. In this regard, it is interesting noting that although the RNA-guided system has not evolved to assess and cleave chromosomal DNA in eukaryotic cell nuclei, it can nonetheless perform relatively well in human cells when compared to TALENs (**Fig. 6**). However, regardless of the nuclease platform ultimately chosen, chromatin impact indexes of two or higher (**Fig. 6**) are relevant for the ultimate performance of genome editing protocols in different experimental settings, including those aiming at bi-allelic or multi-allelic target gene knockouts. Our results complement recent biochemical data demonstrating that nucleosome occupancy obstructs CRISPR/Cas9-mediated DNA cleavage [35]. Interestingly, comparison of targeted DSB formation by TALENs and CRISPR/Cas9 nucleases at different *CCR5* sequences in induced pluripotent stem cells did not show a correlation between indel frequencies and the distribution of DNaseI hypersensitive sites retrieved from the ENCODE database [36]. Whether this finding is specific to the DNaseI hypersensitive site profile selected or to the experimental design used, would need further investigation.

In the present work, we have presented and validated two complementary functional readouts for investigating and comparing in a quantitative manner the role of high-order chromatin structures on genome editing events resulting from the use of sequence-specific designer nucleases. By using these quantitative assays, we have demonstrated that TALENs and CRISPR/Cas9 nucleases are both significantly hindered by closed chromatin in living cells. We conclude that, in addition to the well-established parameters mentioned earlier, the chromatin conformation constitutes yet another determinant of the ultimate efficiencies resulting from using TALENs and CRISPR/Cas9 nucleases. The approaches described in our work should be directly applicable for screening and selecting specific designer nuclease

reagents from different platforms or with different architectures and compositions (e.g. hybrid nucleases, engineered homing endonucleases, Cas9 orthologues, novel Cas9 variants and gene editing tools co-opted from new DNA-targeting systems). In addition, they should also be valuable for assessing the impact of chromatin on different gene editing strategies, including those based on paired 'nickases', truncated guide RNAs and homology-directed gene repair.

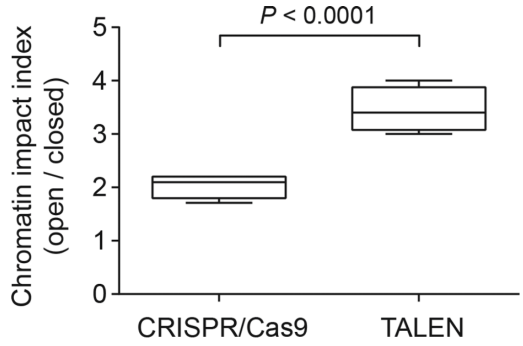


Fig. 6
Cumulative chromatin impact indexes for the TALEN and CRISPR/Cas9 nuclease systems. Boxplot of the ratios between *EGFP* knockout levels measured in the presence versus those determined in the absence of Dox presented in Figures **Figs. 3B**, **D**, **4A** and **5B** (Cas9 data points). *Whiskers*, minimum and maximum. The data corresponding to the TALEN pair with the shortest DNA-binding domains (i.e. TALEN-26-L/TALEN-26-R) was not computed in this analysis to avoid skewing the data. Ten thousand events, each corresponding to a single viable cell, were measured per sample. The *P*-value was determined by two-tailed Student's *t*-test analysis ($P < 0.05$ was considered significant).

Methods

Recombinant DNA

The lentiviral transfer plasmid AD12_pLV.TetO.TLR.TetO (**Supplementary Figure S1**) was constructed by inserting tTR-KRAB binding elements upstream and downstream of the TLR construction [12] in pCVL Traffic Light Reporter 1.1 (SceI target) Ef1a Puro (Addgene plasmid #31482, herein referred to as pLV.TLR) by using standard recombinant DNA techniques. The lentiviral transfer plasmid pLVCT-tTR-KRAB [13] used to assemble LVCT-tTR-KRAB vector particles for the generation of reporter cells HEK.EGFP$^{TetO.KRAB}$, was obtained from Addgene (#11643). Likewise for the TALEN expression plasmids TAL2050 (#39408), TAL2051 (#39409), TAL2072 (#39442), TAL2073 (#39443), TAL2076 (#39446), TAL2077 (#39447), TAL2094 (#39428) and TAL2095 (#39429) encoding, respectively, TALEN-26-L, TALEN-26-R, TALEN-43-L, TALEN-43-R, TALEN-45-L, TALEN-45-R, TALEN-36-L and TALEN-36-R proteins based on the *Xanthomonas axonopodis* pv. *citri* TALE scaffold [14]. The TALEN expression plasmids AR36_pTALEN-GA-L and AR37_pTALEN-GA-R code for, respectively, TALEN-LEGFP and TALEN-REGFP [15]; herein referred to as TALEN-GA-L and TALEN-GA-R, respectively. The TALEN-GA-L and TALEN-GA-R proteins are based on the *Xanthomonas campestris* pv. *armoraciae* TALE scaffold and were custom-designed by GeneArt Gene Synthesis (ThermoFisher Scientific). The expression plasmid hCas9 [16]; herein referred to as pCMV.Cas9, contains a human codon-optimized ORF coding for the *Streptococcus pyogenes* Cas9 nuclease (Addgene plasmid #41815). The gRNA acceptor plasmid S7_pUC.U6.sgRNA.BveI-stuffer has a U6 RNA Pol III promoter for driving gRNA expression and was constructed as follows. The construct pLKO.1-puro.U6.sgRNA.BfuAI.stuffer [17] (Addgene plasmid #50920) was treated with BclI and, subsequently, with the Klenow fragment (both from ThermoFisher Scientific). Next, this vector backbone was dephosphorylated with FastAP (ThermoFisher Scientific) and ligated to a Klenow fragment-blunted 3431-bp *DMD* cDNA fragment harboring four BveI sites. This fragment was isolated after digesting pDysE [18] with EcoRI (ThermoFisher Scientific). Of note, the presence of extra BveI sites aids in achieving complete BveI digestion of the respective gRNA acceptor plasmid. These manoeuvres yielded AA19_pLKO.1-puro.U6.sgRNA.BveI-stuffer (**Supplementary Fig. S2**). Finally, after digesting AA19_pLKO.1-puro.U6.sgRNA.BveI-stuffer with BveI (ThermoFisher Scientific) and EcoRI, the resulting 3822-bp insert was ligated to a 2676-bp fragment obtained by treating cloning vector pUCBM21 (Boehringer Mannheim) with HincII (ThermoFisher Scientific) and EcoRI. This cloning step led to the generation of gRNA acceptor plasmid S7_pUC.U6.sgRNA.BveI-stuffer (**Supplementary Figure S3**). The expression plasmids coding for gRNA^{TLR}, gRNA^{GFP2}, gRNA^{GFP3}, gRNA^{GFP4}, gRNA^{GFP5}, gRNA^{GFP6} and gRNA^{GFP7} were assembled by inserting annealed oligonucleotide pairs 5′-ACCGGTGAGCTCTTATTTGCGTA-3′/5′-AAACTACGCAAATAAGAGCTCAC-3′, 5′-ACCGCTGCCGTCCTCGATGTTG-3′/5′-AAACCAACATCGAGGACGGCAG-3′, 5′-ACCGCCGTCCTCGATGTTGTGG-3′/5′-AAACCCACAACATCGAGGACGG-3, 5′-ACCGGGCACGGGCAGCTTGCCGG-3′/ 5′-AAACCCGGCAAGCTGCCCGTGCC-3′, 5-ACCGTCGCCCTCGAACTTCACCT-3′/ 5′-AAACAGGTGAAGTTCGAGGGCGA-3′,

5′-ACCGTAGGTCAGGGTGGTCACGA-3′/ 5′-AAACTCGTGACCACCCTGACCTA-3′ and 5′-ACCGGCGAGGGCGATGCCACCTA-3′/ 5′-AAACTAGGTGGCATCGCCCTCGC-3′ into BveI-digested S7_pUC.U6.sgRNA.Bvel-stuffer, respectively.

The *EGFP* target sequences for $gRNA^{GFP4}$, $gRNA^{GFP5}$, $gRNA^{GFP6}$ were described previously as NGG site 1, NGG site 2 and NGG site 3, respectively [19]. The target sequence for gRNAGFP7 was also described before [20]. The construct gRNA_GFP_T2 [17] expressing the herein named gRNAGFP1 was obtained from Addgene (plasmid #41820). Plasmid gRNA_Cloning Vector expressing the herein called $gRNA^{Empty}$ (Addgene plasmid #41824) and plasmid AM51_ pUC.U6.$gRNA^{NT}$ encoding, respectively, no gRNA and an irrelevant, non-targeting, gRNA, were used to serve as negative controls. The AM51_ pUC.U6.$gRNA^{NT}$ construct was generated by cloning into BveI-digested S7_pUC.U6.sgRNA.Bvel-stuffer the annealed oligonucleotide pair 5′-ACCGGTGAGCTCTTATTTGCGTAGCTAGCTGAC-3′/ 5′-AAACGTCAGCTAGCTACG-CAAATAAGAGCTCAC-3′.

The generation of the isogenic set of expression plasmids encoding wild-type Cas9, SpCas9-HF1 [19] and eSpCas9(1.1) [21] was done as follows. The plasmid C55_pU.CAG.hrGFP.rBGpA (**Supplementary Fig. S4**) was first digested with NotI and HincII, to remove the *hrGFP* ORF, after which the 5019-bp backbone fragment was extracted from the agarose gel. Next, the backbone was blunted and dephosphorylated by using the Klenow Fragment of the E. coli DNA polymerase I and FastAP, respectively (both from Thermo Fisher Scientific). The constructs VP12 (Addgene plasmid #72247) [19] and eSpCas9(1.1) (Addgene plasmid #71814) [21] encode the high-specificity Cas9 variants SpCas9-HF1 and eSpCas9(1.1), respectively. The ORFs coding for wild-type Cas9, SpCas9-HF1 and eSpCas9(1.1) were isolated from agarose gel after double digesting pCMV.Cas9 [16], VP12 and eSpCas9(1.1) with XbaI/MssI, NotI/MssI and AseI/EcoRI, respectively. The insert fragments corresponding to Cas9, SpCas9-HF1 and eSpCas9(1.1) were then blunted and ligated to the aforementioned plasmid backbone yielding expression plasmids AV62_pCAG.Cas9.rBGpA, AV64_pCAG.SpCas9-HF1.rBGpA and AW01_pCAG.eSpCas9(1.1).rBGpA, respectively.

Cells

HEK293T cells (American Type Culture Collection) and HEK293T cell-derived clone HEK.$EGFP^{TetO.KRAB}$ were cultured as indicated elsewhere [22]. $HER.TLR^{TetO.KRAB}$ cells and their control *TetO*-negative counterparts $HER.TLR^{KRAB}$ cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS), 10 mM $MgCl_2$ and hygromycin B (ThermoFisher Scientific) at a final concentration of 45 μg/ml. These cells were stably transfected with a tTR-KRAB expression plasmid based on the pcDNA3.1/Hygro(+) backbone (Life Technologies). The various cell types were kept at 37°C in a humidified-air 10% CO_2 atmosphere.

Lentiviral vectors

The generation of lentiviral vectors LVCT-tTR-KRAB, LV.TetO.TLR.TetO and LV.TLR was done by transfecting, respectively, pLVCT-tTR-KRAB, pLV.TetO.TLR.TetO and pLV.TLR into HEK293T cells together with packaging construct psPAX2 (Addgene #12260) and pseudotyp-

ing plasmid pLP/VSVG (Life Technologies), essentially as described in detail elsewhere [23,24].

Experimental models

The single cell-derived clone HEK.EGFP$^{TetO.KRAB}$, harboring a tTR-KRAB-regulated *EGFP* allele, was generated by transducing HEK293T cells with LVCT-tTR-KRAB. In brief, 5×10^4 cells were seeded in wells of 24-well plates (Greiner Bio-One). After 24 h, the cells were exposed to 1, 0.5, 0.25, 0.125 and 0.062 Hela cell-transducing units (HTU)/cell of LVCT-tTR-KRAB. At 48 h post-transduction, the cells exposed to each vector multiplicity of infection (MOI) were split into wells of 6-well plates (Greiner Bio-One). At 6 days after transduction, these cells were either kept in regular culture medium or in culture medium supplemented with 500 ng/ml doxycycline (Dox). Three days later EGFP-directed flow cytometry analysis was carried out. The cell culture transduced with an MOI of 0.125 HTU/cell contained 6.9% and 0.34% of EGFP-positive cells in the presence and absence of Dox, respectively. This cell culture was selected for isolating LVCT-tTR-KRAB-transduced clones by adding 200 µl of a cell suspension with a concentration of 1.5 cells/ml into wells of 96-well plates (Greiner Bio-One). To increase the cloning efficiency, the culture medium was supplemented with 50 µM of α-thioglycerol and 20 nM of bathocuprione disulphonate (both from Sigma-Aldrich) [25]. Analysis of Dox-dependent epigenetic regulation of target sequences in individual LVCT-tTR-KRAB-transduced clones by flow cytometry led to the selection of reporter cell clone HEK.EGFP$^{TetO.KRAB}$, which was kept in the presence of Dox (200 ng/ml) where indicated. The HER.TLR$^{TetO.KRAB}$ and the control HER.TLR^{KRAB} reporter cells were generated by transducing HER clone A2 with LV.TetO.TLR.TetO and LV.TLR particles, respectively. Clone A2 is derived from PER.C6 cells [26] and expresses, in a constitutive fashion, tTR-KRAB from a stably integrated pcDNA3.1/Hygro(+) backbone (Life Technologies). In brief, 2×10^5 A2 cells were seeded in wells of 24-well plates (Greiner Bio-One). One day later, 4-fold serial dilutions of LV.TetO.TLR.TetO and LV.TLR clarified supernatant were applied onto the cells. At 48 h post-transduction, the cells were transferred to wells of 6-well plates. At approximately 2 weeks post-transduction, hygromycin (45 µg/ml) and Dox (200 ng/ml) were added into LV.TetO.TLR.TetO transduced cells. The addition of Dox allowed one week later the start of cell selection with puromycin at a final concentration of 1 µg/ml. LV.TLR transduced cells were exposed to hygromycin and puromycin. After 2 weeks in selection medium, puromycin-resistant polyclonal cell populations were expanded for carrying out transfection experiments and, where indicated, kept in the presence of Dox (500 ng/ml).

DNA transfections

Transfections were started by mixing each of the relevant plasmids and 1 mg/ml polyethyleneimine (PEI, Polysciences) in 50 µl of a 150 mM NaCl solution (**Supplementary Tables S1–S4**). These transfections were done in cells cultured in the absence or in the presence of Dox. In addition, for determining gene delivery efficiencies, all transfection mixtures contained a reporter-encoding plasmid. After around 10 s of vigorous vortexing, the DNA-PEI complexes were incubated at room temperature for 15 min and were subsequently directly applied into the medium of the target cells that had been seeded one day before. After 6–8 h,

the transfection mixtures were removed and fresh regular culture medium was added onto the transfected cells. At 3 days post-transfection, reporter-directed flow cytometry was carried out to establish the frequency of transfected cells in the various cultures. Subsequently, the cells were sub-cultured about every 3–4 days. In order to activate target gene expression, cultures of HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$ cells were exposed to Dox at 10 and 14 days post-transfection, respectively. Reporter-directed flow cytometry was performed on HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$ cells 7 and 10 days later, respectively. Where indicated, *TetO*-negative control HER.TLR^{KRAB} cells were subjected to the same conditions as those applied to their HER.TLR$^{TetO.KRAB}$ counterparts.

Flow cytometry

The quantification of reporter-positive cells was done by a BD LSR II flow cytometer (BD Biosciences). The results were analyzed with the aid of BD FACSDiva 6.1.3 software (BD Biosciences) or FlowJo 7.2.2 software (Tree Star). Mock-transfected target cells were used to set background fluorescence. At least 10,000 viable single cells were analyzed per sample.

Chromatin immunoprecipitation (ChIP) and qPCR

The HER.TLR$^{TetO.KRAB}$ cells were cultured in the presence or in the absence of Dox (500 ng/ml) for 11 days, after which a cell fixation protocol available at https://www.activemotif.com, was applied. In brief, 2 ml of a freshly prepared formaldehyde solution was added into the cell culture medium. The culture flasks were subsequently agitated for exactly 15 min at room temperature. Next, 1.1 ml of 2.5 M glycine was applied to stop the fixation process. After a 5-min incubation period at room temperature, the cells were scraped and transferred to a 50-ml tube. The collected cells were subjected to two cycles of centrifugation at 800 ×g for 10 min. After the first cycle, the cells were resuspended in 10 ml of pre-chilled Igepal solution (0.05% Igepal in phosphate buffered saline (PBS)), whilst after the second cycle a 10 ml 0.05% Igepal solution mixed with 100 µl phenylmethanesulfonyl fluoride (Sigma P-7626, 100 mM in ethanol), was used instead. A final centrifugation at 800 ×g for 10 min was carried out to harvest the fixed cell pellets. Subsequently, the ChIP-qPCR assays were performed on 30 µg of cross-linked chromatin according to the HistonePath™ ChIP-qPCR protocol (Active Motif). The ChIP-validated antibodies H3 pan-acetyl (Active Motif, cat # 39139) and H3K9me3 (Active Motif, cat # 39161) were used for the ChIP. Next, qPCR amplifications with primer pairs (**Supplementary Table S5**) targeting six different *TLR^{TetO}* regions (i.e. *SFFV* promoter, *EF1α* promoter, *puromycin resistance* gene, *mCherry* sequence, 5′ and 3′ *EGFP* gene segments) were performed. Additional primer pairs were used for the quality control of the ChIP-qPCR assays (**Supplementary Table S5**).

The qPCR amplifications were performed with the aid of a Bio-Rad CFX Connect Real-time detection system running the CFX Manager™ software. The qPCR amplifications were carried out in triplicate for each sample and primer pair. The qPCR mixtures consisted of 12.5 ng of input DNA, 1x iQ SYBR Green Supermix (Bio-Rad) and 250 nM of each primer in a total volume of 20 µl. The qPCR program started with a 2-min incubation period at 95°C. This was

followed by 40 cycles consisting of 15 s at 95°C, 20 s at 58°C and 20 s at 72°C. Next, the samples were sequentially incubated at 95°C and 55°C for 1 min. The melting curves were derived by increasing the temperature from 55°C to 95°C with a rate of 0.5°C for every 10 s. The binding events detected per 1000 cells were calculated based on chromatin input amounts, final ChIP volumes and primer efficiencies. Finally, the data were normalized using the algorithm available at https://www.activemotif.com.

Statistical analysis

The comparison of data sets resulting from a minimum of three independent experiments were analyzed by using the GraphPad Prism 6 software package and monitored for significance by applying two-tailed Student's *t*-tests with $P < 0.05$ considered significant. The data sets derived from two independent experiments done in biological replicate in HER.TLR^{KRAB} were analyzed by using the IBM SPSS Statistics 23 software package and monitored for significance by employing a three-way ANOVA test with $P < 0.05$ considered significant. One-way ANOVA combined with Bonferroni tests were used for the statistical analysis of the data corresponding to the screening of the high-specificity Cas9 variants. $P < 0.05$ was considered significant. The IBM SPSS Statistics 23 software package was used in this analysis.

Acknowledgments

The authors thank Ron Wolterbeek (Department of Medical Statistics and Bioinformatics, Leiden University Medical Center) for his help with the statistical analysis using the IBM SPSS Statistics 23 software package. Xiaoyu Chen is the recipient of a Ph.D. research grant from the China Scholarship Council-Leiden University Joint Scholarship Programme.

Supplementary Information

Supplementary information can be found at doi: 10.1093/nar/gkw524.

Funding

Dutch Prinses Beatrix Spierfonds [W.OR11–18]; French AFMTéléthon [16621]. Funding for open access charge: Prinses Beatrix Spierfonds [W.OR11–18].

Conflict of interest statement. None declared.

Reference

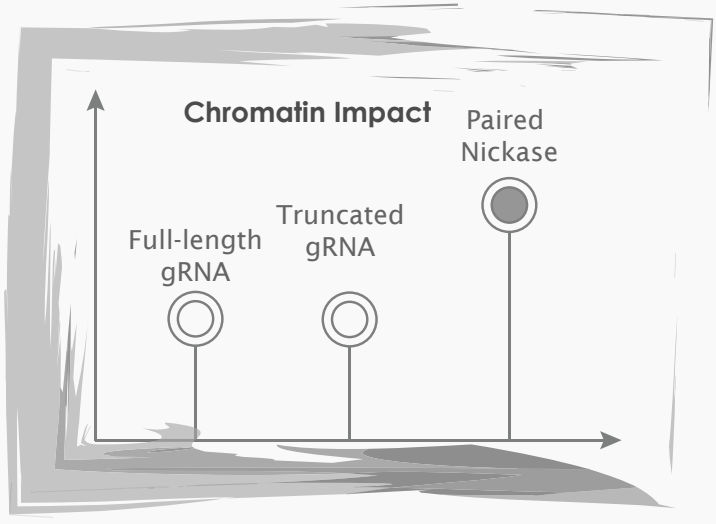
1. Kim H., Kim J.-S. A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. 2014;15:321–334.
2. Maggio I., Gonçalves M.A.F.V. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends. Biotechnol. 2015;33:280–291.
3. Chylinski K., Makarova K.S., Charpentier E., Koonin E.V. Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res. 2014;42:6091–6105.
4. Anders C., Niewoehner O., Duerst A., Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014;513:569–573.
5. Chen S., Oikonomou G., Chiu C.N., Niles B.J., Liu J., Lee D.A., Antoshechkin I., Prober D.A. A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. Nucleic Acids Res. 2013;41:2769–2778.

6. Valton J., Dupuy A., Daboussi F., Thomas S., Maréchal A., Macmaster R., Melliand K., Juillerat A., Duchateau P. Overcoming Transcription Activator-like Effector (TALE) DNA binding domain sensitivity to cytosine methylation. J. Biol. Chem. 2012;287:38427–38432.
7. Wu X., Kriz A.J., Sharp P.A. Target specificity of the CRISPR-Cas9 system. Quant. Biol. 2014;2:59–70.
8. Kuscu C., Arslan S., Singh R., Thorpe J., Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat. Biotechnol. 2014;32:677–683.
9. O'Geen H., Henry I.M., Bhakta M.S., Meckler J.F., Segal D.J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 2015;43:3389–3404.
10. Polstein L.R., Perez-Pinera P., Kocak D.D., Vockley C.M., Bledsoe P., Song L., Safi A., Crawford G.E., Reddy T.E., Gersbach C.A. Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. Genome Res. 2015;25:1158–1169. [
11. Wu X., Scott D.A., Kriz A.J., Chiu A.C., Hsu P.D., Dadon D.B., Cheng A.W., Trevino A.E., Konermann S., Chen S., et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 2014;32:670–676.
12. Certo M.T., Ryu B.Y., Annis J.E., Garibov M., Jarjour J., Rawlings D.J., Scharenberg A.M. Tracking genome engineering outcome at individual DNA breakpoints. Nat. Methods. 2011;8:671–676.
13. Szulc J., Wiznerowicz M., Sauvain M.-O., Trono D., Aebischer P. A versatile tool for conditional gene expression and knockdown. Nat. Methods. 2006;3:109–116.
14. Reyon D., Tsai S.Q., Khayter C., Foden J.A., Sander J.D., Joung J.K. FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. 2012;30:460–465.
15. 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. 2014;11:1051–1057.
16. 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. 2013;339:823–826.
17. Kearns N.A., Genga R.M.J., Enuameh M.S., Garber M., Wolfe S.A., Maehr R. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development. 2013;141:219–223.
18. Chapdelaine P., Moisset P.-A., Campeau P., Asselin I., Vilquin J.-T., Tremblay J.P. Functional EGFP–dystrophin fusion proteins for gene therapy vector development. Protein Eng. 2000;13:611–615.
19. Kleinstiver B.P., Pattanayak V., Prew M.S., Tsai S.Q., Nguyen N.T., Zheng Z., Joung J.K. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529:490–495.
20. Fu Y., Sander J.D., Reyon D., Cascio V.M., Joung J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 2014;32:279–284.
21. Slaymaker I.M., Gao L., Zetsche B., Scott D.A., Yan W.X., Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351:84–88.
22. Maggio I., Stefanucci L., Janssen J.M., Liu J., Chen X., Mouly V., Goncalves M.A. Selection-free gene repair after adenoviral vector transduction of designer nucleases: rescue of dystrophin synthesis in DMD muscle cell populations. Nucleic Acids Res. 2016;44:1449–1470.
23. Pelascini L.P., Goncalves M.A. Lentiviral vectors encoding zinc-finger nucleases specific for the model target locus HPRT1. Methods. Mol. Biol. 2014;1114:181–199.
24. Pelascini L.P., Janssen J.M., Goncalves M.A. Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. Hum. Gene Ther. 2013;24:78–96.

25. Brielmeier M., Béchet J.M., Falk M.H., Pawlita M., Polack A., Bornkamm G.W. Improving stable transfection efficiency: antioxidants dramatically improve the outgrowth of clones under dominant marker selection. Nucleic Acids Res. 1998;26:2082–2085.
26. 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. New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. Hum. Gene Ther. 1998;9:1909–1917.
27. Lechner M.S., Begg G.E., Speicher D.W., Rauscher F.J. Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain–KAP-1 corepressor interaction is essential. Mol. Cell. Biol. 2000;20:6449–6465.
28. Nielsen A.L., Ortiz J.A., You J., Oulad-Abdelghani M., Khechumian R., Gansmuller A., Chambon P., Losson R. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J. 1999;18:6385–6395.
29. Ayyanathan K., Lechner M.S., Bell P., Maul G.G., Schultz D.C., Yamada Y., Tanaka K., Torigoe K., Rauscher F.J. Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev. 2003;17:1855–1869.
30. Groner A.C., Meylan S., Ciuffi A., Zangger N., Ambrosini G., Denervaud N., Bucher P., Trono D. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet. 2010;6:e1000869.
31. Becker J.S., Nicetto D., Zaret K.S. H3K9me3-Dependent Heterochromatin: Barrier to cell fate changes. Trends Genet. 2016;32:29–41.
32. Mahfouz M.M., Li L., Shamimuzzaman M., Wibowo A., Fang X., Zhu J.K. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc. Natl. Acad. Sci. U.S.A. 2011;108:2623–2628.
33. 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., et al. A TALE nuclease architecture for efficient genome editing. Nat. Biotech. 2011;29:143–148.
34. Nishimasu H., Ran F.A., Hsu P.D., Konermann S., Shehata S.I., Dohmae N., Ishitani R., Zhang F., Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;156:935–949.
35. Horlbeck M.A., Witkowsky L.B., Guglielmi B., Replogle J.M., Gilbert L.A., Villalta J.E., Torigoe S.E., Tjian R., Weissman J.S. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. ELife. 2016;5:e12677.
36. 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. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 2013;41:9049–9061.

Chapter 4

The **Chromatin** Structure **Differentially** Impacts **High-specificity** CRISPR-Cas9 Nuclease Strategies.



Molecular Therapy-Nucleic Acids 8:558–563 (2017).
Chen X, Liu J, Janssen JM, Gonçalves MA.

Genome editing technologies based on RNA-guided nucleases (RGNs) derived from prokaryotic type II CRISPR-Cas9 adaptive immune systems, such as that from *Streptococcus pyogenes*[1, 2] and, more recently, *Staphylococcus aureus*,[3, 4] are becoming increasingly pervasive in both basic and applied research.[5, 6] RGNs are ribonucleoprotein complexes whose sequence-specific guide RNA (gRNA) moieties address a Cas9 nuclease to a DNA target site (**Figure S1A**). Base pairing between the 5′-terminal nucleotides of the gRNA (spacer) and DNA sequences connected to a protospacer-adjacent motif (PAM), triggers Cas9-mediated double-stranded DNA break (DSB) formation.[5, 6] RGNs based on *S. pyogenes Cas9* (SpCas9) and on *S. aureus* Cas9 (SaCas9) typically have spacers with a length of 20 and 21–24 nucleotides, respectively. The PAM of SpCas9 is NGG, while that of SaCas9 is NNGRRT.[3, 6] The repair of RGN-induced targeted DNA lesions by non-homologous end joining (NHEJ) or homologous recombination can result in either the deletion or addition of genetic information in cells from virtually any organism.[5, 6]

Despite their ease of use and broad applicability, a major limitation of conventional RGNs is that of off-target DNA cleavage. Indeed, it has been demonstrated that up to five gRNA-DNA mismatches can be tolerated, especially when located most distally to the PAM.[7] Such mismatches can conceivably result in several hundreds of DSBs across the genome.[7] Therefore, substantial efforts have been made in recent years to minimize RGN off-target activities. These efforts resulted in a set of improved genome editing strategies, of which preeminent examples include: (1) RGNs harboring 5′-truncated gRNAs (tru-gRNAs),[8] (2) rationally designed high-specificity SpCas9 variants,[9, 10] (3) offset RGN pairs containing nicking SpCas9 mutants,[11, 12] and (4) repurposing of RGNs with longer PAMs, such as those from the *S. aureus* CRISPR/Cas9 system.[3, 4] An overview of these approaches is presented in **Figure S1**.

Previous studies have shown a preferential interaction between SpCas9-based RGNs and genomic regions with euchromatic signatures.[13, 14, 15, 16, 17] Recently, by using cellular models based on the conditional recruitment of epigenetic remodeling complexes to isogenic target sequences, our laboratory, and that of others, have demonstrated that the activity of conventional programmable nucleases, including RGNs, can be significantly hindered by compact heterochromatin in living cells.[18, 19] However, so far, the extent to which high-specificity CRISPR/Cas9 nuclease strategies are affected by higher-order chromatin topologies remains to be determined.

Hence, here, we sought to investigate and compare the impact that epigenetically regulated three-dimensional chromatin "conformers" have on the performance of the aforementioned high-specificity genome editing principles (**Figure S1**). In these experiments, we deployed complementary loss-of-function and gain-of-function cellular systems in which the euchromatic and heterochromatic statuses of isogenic target sequences are controlled by doxycycline (Dox).[18] These systems, based

on clonal HEK.EGFP$^{TetO.KRAB}$ and polyclonal HER.TLR$^{TetO.KRAB}$ cell lines, permit the measurement of the frequencies of targeted DSBs made by different programmable nucleases through the quantification of EGFP− and mCherry+ cells,[18, 20] respectively, generated after NHEJ-mediated DSB repair (for details, see **Figure 1A**). In brief, reporter cells cultured without Dox (heterochromatic target sites; high H3K9me3/low H3Ac) or with Dox (euchromatic target sites; high H3Ac/low H3K9me3) are exposed to different programmable nuclease combinations.[18] After the action of the programmable nucleases takes place, all cultures receive Dox to allow for transgene expression and quantification of gene editing events (**Figure 1A**).

We started by comparing site-specific DSB formation by RGNs containing full-length gRNAs (FL-gRNAs) or Tru-gRNAs at euchromatic ("open") versus heterochromatic ("closed") target sequences. The former gRNAs have canonical, 20-mer spacers; the latter display shorter, mostly 18-mer to 17-mer spacers.[8] The reduced DNA-binding energies of Tru-gRNAs is thought to cause the preferential binding of the respective SpCas9 partner to fully complementary target DNA (**Figure S1B**).[8] Gene editing experiments were initiated by transfecting HEK.EGFP$^{TetO.KRAB}$ cells, cultured in the presence of Dox or on its absence, with plasmids encoding sets of SpCas9:FL-gRNA or SpCas9:Tru-gRNA complexes targeting five different positions along the *EGFP* ORF (**Figure S2**). After the action of the various RGN complexes had taken place, Dox was added to the different cell cultures for flow cytometric quantification of targeted gene knockout levels. We found that, independently of their lengths, the various RGNs had similar chromatin impact indexes (**Figure 1B**), as defined by the ratios between the frequencies of DSB formation at euchromatic versus heterochromatic target sites (**Figure S3**). Hence, despite their predicted lower DNA-binding energies, derived from a reduced Watson and Crick base-pairing potential, RGNs with Tru-gRNAs were hindered by heterochromatin to approximately the same extent as those harboring standard FL-gRNAs.

The recognition of the PAM by the PID domain of SpCas9 is the first event leading to targeted DNA cleavage.[21] After this initial genomic DNA interrogation, local double helix melting permits the nucleation of gRNA-DNA hybridization and subsequent R-loop expansion along a PAM-proximal to PAM-distal direction (i.e., 3′ to 5′). Finally, full-length heteroduplex formation between unwound DNA and gRNA sequences triggers phosphodiester bond hydrolysis on both strands through a concerted, PAM-dependent, allosteric activation of SpCas9's HNH and RuvC nuclease domains.[21] Based on this series of events, our data (**Figure 1B**) indicate that, once a catalytically competent RGN complex manages to engage a heterochromatic PAM sequence, the epigenetic barrier has, for the most part, been overcome, with the length of the gRNA posing no significant limitations to the subsequent aforementioned downstream processes. As corollary, the ultimate activity of SpCas9:gRNA complexes at specific, PAM-defined, target sites seems to be primarily determined by

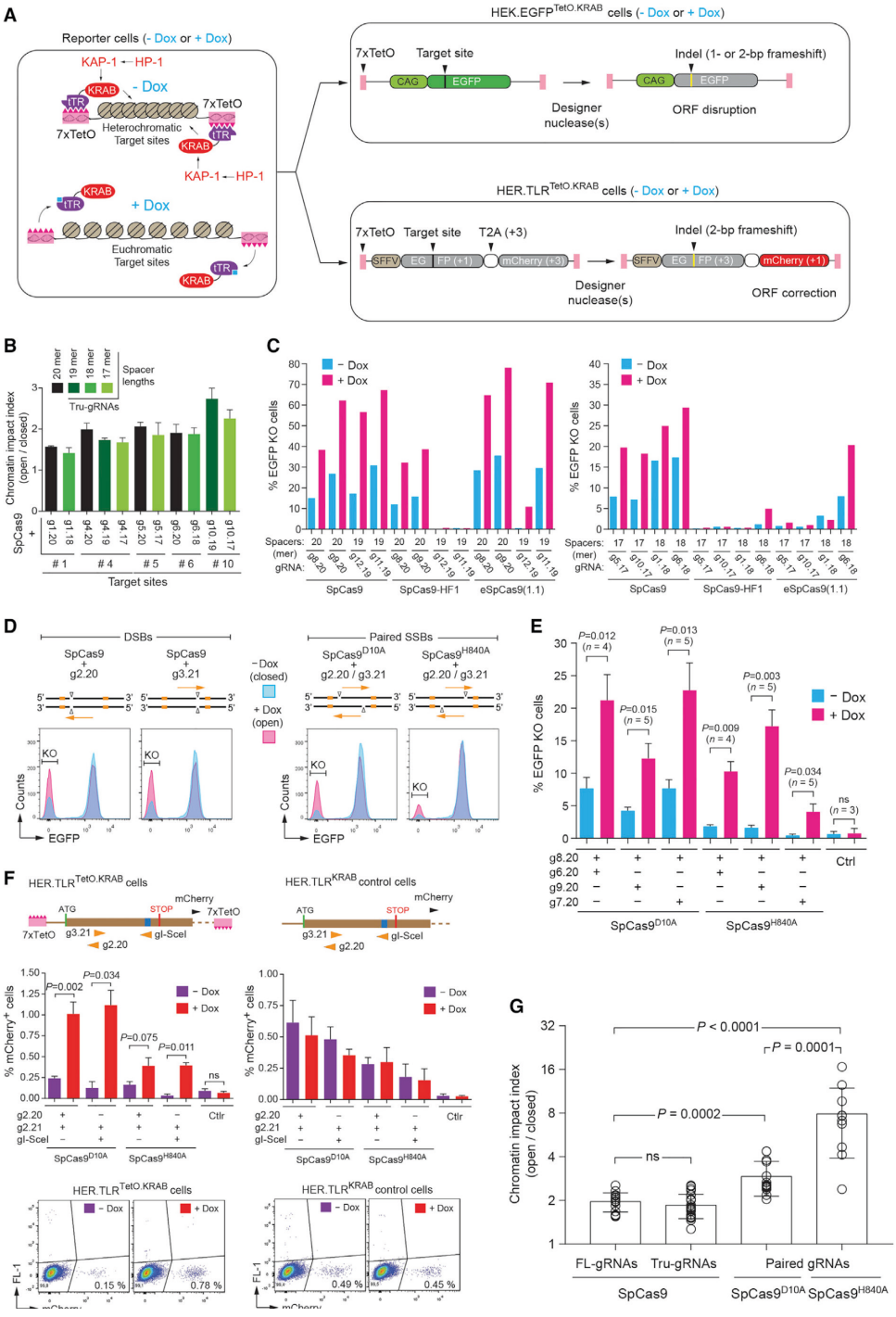

A
Reporter cells (- Dox or + Dox)
KAP-1 HP-1
KRAB
- Dox
tTR
7xTetO
Heterochromatic Target sites
+ Dox
Euchromatic Target sites
HEK.EGFP TetO.KRAB cells (- Dox or + Dox)
7xTetO
Target site
CAG
EGFP
Indel (1- or 2-bp frameshift)
Designer nuclease(s)
ORF disruption
HER.TLR TetO.KRAB cells (- Dox or + Dox)
T2A (+3)
SFFV
EG
FP (+1)
mCherry (+3)
FP (+3)
mCherry (+1)
ORF correction
B
Chromatin impact index (open / closed)
20 mer
19 mer
18 mer
17 mer
Spacer lengths
Tru-gRNAs
SpCas9
Target sites
C
- Dox
+ Dox
% EGFP KO cells
Spacers: (mer)
gRNA:
SpCas9
SpCas9-HF1
eSpCas9(1.1)
D
DSBs
SpCas9 + g2.20
SpCas9 + g3.21
Paired SSBs
SpCas9D10A + g2.20 / g3.21
SpCas9H840A + g2.20 / g3.21
- Dox (closed)
+ Dox (open)
KO
Counts
EGFP
E
% EGFP KO cells
P=0.012 (n = 4)
P=0.015 (n = 5)
P=0.013 (n = 5)
P=0.009 (n = 4)
P=0.003 (n = 5)
P=0.034 (n = 5)
ns (n = 3)
g8.20
g6.20
g9.20
g7.20
Ctrl
SpCas9D10A
SpCas9H840A
F
HER.TLR TetO.KRAB cells
HER.TLRKRAB control cells
mCherry
ATG
STOP
7xTetO
g3.21
g2.20
gI-SceI
% mCherry+ cells
P=0.002
P=0.034
P=0.075
P=0.011
ns
g2.21
0.15 %
0.78 %
0.49 %
0.45 %
FL-1
G
Chromatin impact index (open / closed)
P < 0.0001
P = 0.0001
P = 0.0002
ns
FL-gRNAs
Tru-gRNAs
Paired gRNAs
SpCas9
SpCas9D10A
SpCas9H840A

Figure 1 Assessing the effect of alternative chromatin topologies on high-specificity CRISPR/Cas9 nucleases based on SpCas9 proteins (**A**) Cellular reporter systems for tracking programmable nuclease-induced DSB formation at isogenic target sequences possessing different chromatin conformations. Left panel: the binding of tTR-KRAB fusion proteins to *TetO* elements triggers heterochromatin spread after the recruitment of endogenous epigenetic remodeling complexes composed of, among others, KAP1 and HP1. The addition of doxycycline (Dox) alters the tTR-KRAB conformation, preventing its interaction with the *TetO* elements, resulting in the transition of the associated DNA sequences from a compact/heterochromatic into a relaxed/euchromatic state. Top right panel: schematics of the HEK.EGFP$^{TetO.KRAB}$ loss-of-function system. The tTR-KRAB-expressing human embryonic kidney cells HEK.EGFP$^{TetO.KRAB}$ harbor a *TetO*-flanked *EGFP* expression unit. In this construct, an in-frame EGFP reporter can be placed out of frame after indel formation by NHEJ-mediated repair of site-specific DSBs. Thus, non-fluorescent cells report sequence-specific nuclease activity. Bottom right panel: diagram of the HER.TLR$^{TetO.KRAB}$ gain-of-function system. The tTR-KRAB-expressing human embryonic retinoblasts HER.TLR$^{TetO.KRAB}$ contain a Traffic Light Reporter (*TLR*)[20] flanked by *TetO* elements (*TLR^{TetO}*).[18] In this construct, an out-of-frame *mCherry* ORF linked to a *T2A* sequence and an *EGFP* ORF with a disrupting I-SceI recognition site, can be placed in-frame after indel formation by NHEJ-mediated repair of targeted DSBs. Hence, in this case, the resulting red fluorescent cells are those reporting sequence-specific nuclease activity. (**B**) Chromatin impact indexes of RGNs with full-length and truncated gRNAs. HEK.EGFP$^{TetO.KRAB}$ cells, treated and not treated with Dox, were exposed to SpCas9 together with the indicated gRNAs. The chromatin impact index for each RGN was determined by dividing the *EGFP* knockout levels measured at euchromatic by those gauged at heterochromatic conditions (i.e., +Dox and −Dox, respectively). Error bars indicated mean ± SD (n = 3 independent experiments done on different days). (**C**) Testing the impact of chromatin on RGNs consisting of high-specificity Cas9 variants and Tru-gRNAs. HEK.EGFP$^{TetO.KRAB}$ cells, incubated in the presence and in the absence of Dox, received the indicated nuclease-gRNA pairs. Targeted mutagenesis levels at euchromatic (+Dox) and heterochromatic (−Dox) target sequences were quantified by EGFP-directed flow cytometry. The bars in both graphs correspond to an individual experiment using RGNs with FL-gRNAs (n = 2) or Tru-gRNAs (n = 6) (**D**) Validation of HEK.EGFP$^{TetO.KRAB}$ cells for assessing paired RGN "nickases". HEK.EGFP$^{TetO.KRAB}$ cells, treated and not treated with Dox, were exposed to SpCas9 and single gRNAs (DSBs) or SpCas9 nicking mutants together with two partially overlapping gRNAs (paired SSBs). Cell fractions with *EGFP* knockout (KO) alleles are indicated in the flow cytometry histograms. (**E**) Screening dual RGN "nickase" activities at euchromatic versus heterochromatic target sequences. HEK.EGFP$^{TetO.KRAB}$ cells, treated and not treated with Dox, were exposed to the specified dual RGN "nickases". The frequencies of targeted mutagenesis under both Dox regimens were determined by EGFP-directed flow cytometry. Error bars indicate mean ± SEM, corresponding to independent experiments (n) performed on different days. (**F**) Testing of dual RGN nicking complexes on the gain-of-function reporter system. HER.TLR$^{TetO.KRAB}$ and control HER.TLR^{KRAB} cells, incubated in the presence and absence of Dox, received the indicated dual RGN "nickases". The frequencies of DSB-induced ORF correction under both Dox regimens were determined by mCherry-directed flow cytometry. Error bars indicate mean ± SEM. A minimum of two and a maximum of four independent experiments were carried out. Representative flow cytometry dot plots of these experiments are presented below each graph. (**G**) Comparison of chromatin impact indexes between conventional and high-specificity RGN complexes. Scatterplot with bar, gathers the datasets obtained from gene editing experiments carried out in HEK.EGFP$^{TetO.KRAB}$ cells.

the degree of chromatin accessibility rather than the extent of gRNA-DNA hybridization at those sites. These results *bode well* for using 18-mer and 17-mer Tru-gRNAs for achieving strict target DNA cleavage[8] and possibly deploying <16-mer Tru-gRNAs, which render RGNs catalytically inert, for multiplexing purposes. Indeed, the combination of these shortened, <16-mer Tru-gRNAs together with SpCas9 proteins fused to heterologous domains can be applied in a variety of orthogonal contexts,

such as those involving combinatorial editing and transcriptional modulation of distinct loci in individual cells.[22, 23]

In a previous study, our laboratory demonstrated that RGNs consisting of FL-gRNAs and SpCas9-HF1 or FL-gRNAs and eSpCas9(1.1) are hindered by heterochromatin[18]. Combining Tru-gRNAs with high-specificity Cas9 variant SpCas9-HF1 (Kleinstiver et al. [9]) or eSpCas9(1.1)[10] offers the prospect for further minimizing off-target DSB formation (**Figure S1C**). Previous experiments have, however, indicated that such Tru-gRNA-containing RGNs display reduced on-target activities when compared to those bearing FL-gRNAs.[9, 10] Results presented in **Figure 1C** extend these findings to isogenic target sequences subjected to distinct epigenetic states. In addition, our data indicate that the compatibility between Tru-gRNAs and eSpCas9(1.1) is higher than that between Tru-gRNAs and SpCas9-HF1 in that, when compared to the former, the latter setting yielded lower gene knockout frequencies at each of the tested PAM-defined target sequences. Indeed, combining SpCas9-HF1 and Tru-gRNAs with spacers a single nucleotide shorter than 20-mer FL-gRNAs (i.e., g11.19 and g12.19) sufficed to abrogate RGN activity in cells subjected to both Dox regimens, but this was not so when applying the same Tru-gRNAs together with eSpCas9(1.1) (**Figure 1C**, left graph). Additional gene editing experiments using a panel of <19-mer Tru-gRNAs (i.e., g5.17, g10.17, g1.18 and g6.18) in HEK.EGFP$^{TetO.KRAB}$ cells treated and not treated with Dox confirmed the higher catalytic impairment of SpCas9-HF1:Tru-gRNA complexes over their eSpCas9(1.1):Tru-gRNA counterparts (**Figure 1C**, right graph).

Offset gRNA pairs addressing SpCas9 "nickases" to opposite DNA strands guarantee that DSBs are, for the most part, restricted to the bipartite target site after the local coordinated formation of SSBs (**Figure S1D**). Indeed, specificity gains between 200-fold to >1,500-fold have been reported for this dual RGN nicking strategy.[12] Gene editing experiments with two gRNAs (i.e., g2.20 and g3.21), whose target sequences partially overlap (**Figure S2**), were individually combined with wild-type SpCas9 (controls) or mixed together with SpCas9^{D10A} or with SpCas9^{H840A}. These experiments validated HEK.EGFP$^{TetO.KRAB}$ cells as a readout system for assessing the impact of chromatin on dual RGN "nickases" (**Figure 1D**). Albeit dependent on two different gRNAs, dual RGN "nickases" display a higher theoretical coverage of the genomic landscape when compared to that of conventional RGNs and other dimeric programmable nucleases, such as zinc-finger nucleases.[5] This stems from the fact that dual RGN "nickases" are compatible with a broad range of DNA spacing between the target sites of their individual members (i.e., about –4-bp to +100-bp), which, in turn, increases the chances for locating suitable PAMs. Hence, to exploit this feature, in subsequent gene editing experiments performed in HEK.EGFP$^{TetO.KRAB}$ cells, we focused on using sets of gRNA pairs (i.e. g8.20/g6.20, g8.20/g9.20, and g8.20/g7.20) with non-overlapping target sequences (**Figure S2**). Flow cytometric quantification

of EGFP− cells generated in HEK.EGFP$^{TetO.KRAB}$ cultures after NHEJ-mediated DSB repair revealed that dual RGN "nickases" were significantly impaired at heterochromatic target sites (**Figure 1E**). This was independent of the nicking SpCas9 mutant used and, therefore, independent of the type of single-stranded DNA overhangs generated (**Figure 1E**). Similar results were obtained after flow cytometric quantification of mCherry+ cells in HER.TLR$^{TetO.KRAB}$ cultures subjected to both Dox regimens and exposed to different gRNA pairs (i.e., g2.20/gI-SceI and g2.21/gI-SceI) combined with each of the two nicking SpCas9 mutants (**Figure 1F** and **Figure S4**). Importantly, there were no significant Dox-dependent differences in the frequencies of mCherry+ cells in control, *TetO*−, HER.TLR^{KRAB} cultures subjected to the same experimental conditions that had been applied to HER.TLR$^{TetO.KRAB}$ cultures. Taken together, these experiments show that, in contrast to RGNs containing Tru-gRNAs, dual RGN "nickases" are significantly more affected by heterochromatin than conventional RGNs (**Figure 1G**). We speculate that, when heterochromatinized, the intervening sequences separating the two opposite SSBs made by dual RGN "nickases" might become more resistant to double helix denaturation when compared to their euchromatinized isogenic counterparts. Interestingly, of the two types of dual RGN "nickases," those based on SpCas9^{D10A} are the least hindered by heterochromatin. SpCas9^{D10A} cuts the DNA strand complementary to the gRNA, whereas SpCas9^{H840A} cleaves the non-complementary strand. Moreover, in contrast to SpCas9^{D10A}, SpCas9^{H840A} has 3′ to 5′ exonuclease activity.[2] Whether these or other biochemical traits underlie the observed differential impact of heterochromatin on the activity of these dual RGN "nickases" will be worthy of further investigation.

The relatively small size of the SaCas9 coding sequence (3.3 Kb) permits its incorporation, together with regulatory elements and gRNA expression units, into viral vector particles with limited packaging capacity, such as those of commonly used adeno-associated viral vectors.[3] In addition, SaCas9-based RGNs increase gene editing versatility by permitting orthogonal (epi)genetic manipulations and potentially offer a higher degree of target site specificity owing to their extended spacer and PAM sequences. Recently, a molecular evolution strategy led to the selection of a SaCas9 variant with an expanded targeting range, i.e., SaCas9^{KKH} (PAM = NNNRRT instead of NNGRRT), further increasing the versatility of SaCas9-based RGNs.[4] As of yet, the impact of chromatin on these new gene editing tools has not been studied. Therefore, we next carried out gene editing experiments in HEK.EGFP$^{TetO.KRAB}$ cells using SaCas9:gRNA and SaCas9^{KKH}:gRNA complexes targeting three different *EGFP* sites embedded in euchromatin or heterochromatin (**Figure 2A**). As references, we also targeted each of these target sequences with prototypic SpCas9:gRNA complexes (**Figure 2A**). Results presented in **Figures 2B** and **2C** reveal that the SaCas9 RGNs were clearly hindered by heterochromatin. In fact, in contrast to SpCas9 RGNs, at heterochromatin, SaCas9 RGNs yielded frequencies of EGFP knockout that were at

or only slightly above background levels.

We conclude that the higher-order chromatin environment is an important parameter to take into consideration while selecting and designing the tools and strategies underlying precise genome editing based on high-specificity CRISPR/Cas9 nuclease complexes.

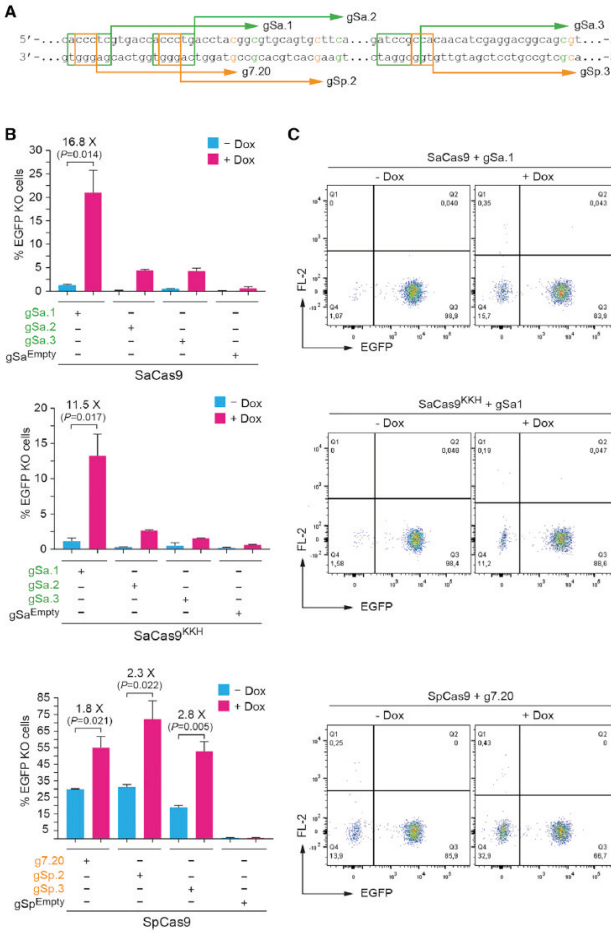


Figure 2. Assessing the effect of alternative chromatin conformations on CRISPR/Cas9 nucleases based on SaCas9 proteins (**A**) Target sequences of RGNs consisting of SaCas9:gRNA or SpCas9:gRNA complexes. The sequences complementary to the *S. aureus* and *S. pyogenes* gRNAs are indicated underneath the green and orange arrows, respectively. The sequences corresponding to the PAMs of *S. aureus* and *S. pyogenes* RGNs are highlighted by green and orange boxes, respectively. (**B**) Testing the effect of chromatin on RGNs harboring SaCas9, SaCas9^{KKH}, or SpCas9. HEK.EGFP$^{TetO.KRAB}$ cells, incubated in the presence and absence of Dox, received the indicated nuclease-gRNA pairs. Targeted mutagenesis levels at euchromatic (+Dox) and heterochromatic (−Dox) target sequences were quantified by EGFP-directed flow cytometry. Error bars correspond to mean ± SD (n = 3 independent experiments performed on different days). (**C**) Representative flow cytometry dot plots of HEK. EGFP$^{TetO.KRAB}$ cells treated with the indicated experimental conditions.

Methods

Cells

The generation of the human embryonic kidney cell-derived line HEK.EGFP$^{TetO.KRAB}$ was detailed elsewhere.[1] These cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Thermofisher Scientific) supplemented with 10% FBS and 10 mM $MgCl_2$. The generation and characterization of the human embryonic retinoblast-derived lines HER.TLR$^{TetO.KRAB}$ and HER.TLR^{KRAB} were equally detailed elsewhere.[1] These cells were maintained in DMEM containing 10% FBS, 10 mM $MgCl_2$ and 45 μg/ml hygromycin B (ThermoFisher Scientific). The cultures of the various tTR-KRAB-expressing reporter cells were kept at 37°C in an humidified-air 10% CO_2 atmosphere.

Recombinant DNA

The previously described isogenic set of constructs AV62_pCAG.Cas9.rBGpA, AV64_pCAG.SpCas9-HF1.rBGpA and AW01_pCAG.eSpCas9(1.1).rBGpA contain human codon-optimized ORFs coding for the Streptococcus pyogenes proteins SpCas9, SpCas9-HF1 and eSpCas9(1.1), respectively.[1] The generation of the isogenic set of expression plasmids BA15_pU.CAG.SaCas9.rBGpA and BA16_pU.CAG.dCas9.rBGpA and BA19_pU.CAG.SaCas9^{KKH}.rBGpA was performed by inserting into the multiple cloning site of C55_pU.CAG.hrGFP.rBGpA (**Figure S6**), DNA fragments containing human codon-optimized ORFs coding for the *Staphylococcus aureus* proteins SaCas9 [3], catalytically inert ("dead") dSaCas9 [3] and SaCas9^{KKH} [4], respectively. The DNA inserts containing the *SaCas9*, *dSaCas9* and *SaCas9^{KKH}* ORFs were isolated from constructs pAAV.CMV.NLS.SaCas9.NLS.3xHA.bGHpA (Addgene plasmid #61592), pAAV.CMV.NLS.dSaCas9.NLS.3xHA.bGHpA (Addgene plasmid #61594) and MSP1830 (Addgene plasmid #70708), respectively.

The plasmid hCas9_D10A,[2] Addgene plasmid #41816, herein named AT42_pSpCas9^{D10A}, expresses the nicking SpCas9^{D10A} variant whose nuclease RuvC-like domain is disrupted by the D10A mutation. The plasmid AT61_pSpCas9^{H840A} (**Figure S5**) expresses the nicking SpCas9^{H840A} variant whose nuclease HNH domain is disabled by the H840A mutation. This expression plasmid was generated as follows. The hCas9 construct,[2] Addgene plasmid #41815, was digested with BstXI (ThermoFisher Scientific) and BmgBI (New England Biolabs), after which the plasmid backbone was purified from an agarose gel by using the JetQuick Gel Extraction Spin Kit (Genomed) according to the manufacturer's instructions. The H840A mutation was introduced by site-directed PCR mutagenesis using as template hCas9 together with the primer pair and cycling conditions indicated in **Supplementary Tables 1** and **2**, respectively. The resulting amplicon was subsequently digested with BstXI and BmgBI, purified from agarose gel as aforementioned and ligated to the isolated plasmid backbone. The resulting construct AT61_pSpCas9^{H840A} was verified by Sanger sequencing. Control constructs gRNA_Cloning Vector (Addgene plasmid #41824) and AM51_pUC.U6.gRNA^{NT}, herein dubbed pgEmpty and pgNT, express no gRNA and an irrelevant non-targeting gRNA. Where indicated, expression plasmids C55_pCAG.hrGFP.rBGpA (**Figure S6**) and AM37_pCMV.DsRedEx2.1.bGHpA (**Fig-**

ure S7) were used as internal controls for determining recombinant DNA delivery efficiencies at 3 days post-transfection.

The construct gRNA_GFP_T2, herein named AT44_pg8.20, was obtained from Addgene plasmid #41820.[2] The *Streptococcus pyogenes* gRNA expressing plasmids generated in this study were made by ligating annealed oligonucleotide pairs into the BveI-digested acceptor construct S07_pUC.U6.sgRNA.BveI-stuffer.[1] The *Staphylococcus aureus* gRNA expressing plasmids generated in this study were made by ligating annealed oligonucleotide pairs into the Esp3I-digested acceptor construct BPK2660 [4] (Addgene plasmid #70709). The oligonucleotides used to clone the various gRNA expression constructs are listed in **Supplementary Table 3**.

Gene editing experiments

Gene editing experiments in HEK.EGFP$^{TetO.KRAB}$, HER.TLR$^{TetO.KRAB}$ and HER.TLR^{KRAB} cells, cultured in the presence or the in absence of doxycycline (Dox), were essentially carried out as detailed elsewhere.[1] In brief, HEK- and HER-based lines were exposed for 7 and 11 days, respectively, to regular medium or to medium supplemented with Dox, after which they were seeded in wells of 24-well plates (Greiner Bio-One). The next day, the cells were transfected according to the schemes listed in **Supplementary Tables 4**-**8**. The transfections were initiated by adding, dropwise, a 1 mg/ml polyethyleneimine solution (PEI, Polysciences) to each of the plasmid mixtures diluted in 150 mM NaCl. Subsequently, the resulting 50-µl transfection mixtures were subjected to vigorous vortexing for 10 sec. The DNA-PEI complexes were then let to be formed for 15 min at room temperature and were subsequently directly added into the culture medium of the target cells. After 6-8 hours, the transfection medium was changed for regular culture medium. At 3 days post-transfection, the transfection efficiencies were measured by reporter-directed flow cytometry. Next, the cells were sub-cultured every 3-4 days. In order to activate target transgene expression, the parallel cultures of HEK- and HER-based lines that had not been initially exposed to Dox receive the drug at 10 and 14 days post-transfection, respectively. Finally, reporter-directed flow cytometry was used to quantify the levels of gene editing in HEK- and HER-based lines after an additional period of 7 and 10 days, respectively.

Flow cytometry

Flow cytometry analyses were done with the aid of a BD LSR II flow cytometer (BD Biosciences). The data were analysed by using the FlowJo 7.2.2 software (Tree Star). Non-transfected cells were used to set the gates for the levels of background fluorescence. A minimum of 10,000 viable single cells were acquired per sample.

Statistical analysis

Data sets resulting from independent experiments carried out in different days were analysed for significance with GraphPad Prism 7 software by applying two-tailed Student's *t*-tests with $P < 0.05$ considered significant.

Acknowledgments

The authors thank Rob Hoeben and Ignazio Maggio (Leiden University Medical Center, Departments of Molecular Cell Biology and Pediatrics, respectively) for their critical reading of the manuscript. This work was partially supported by the Dutch Prinses Beatrix Spierfonds (W.OR11–18) and ProQR Therapeutics (Leiden, the Netherlands). X.C. holds a Ph.D. research grant from the China Scholarship Council-Leiden University Joint Scholarship Programme.

Supplementary Information

Supplementary information can be found at doi: 10.1016/j.omtn.2017.08.005

Conflict of interest statement. None declared.

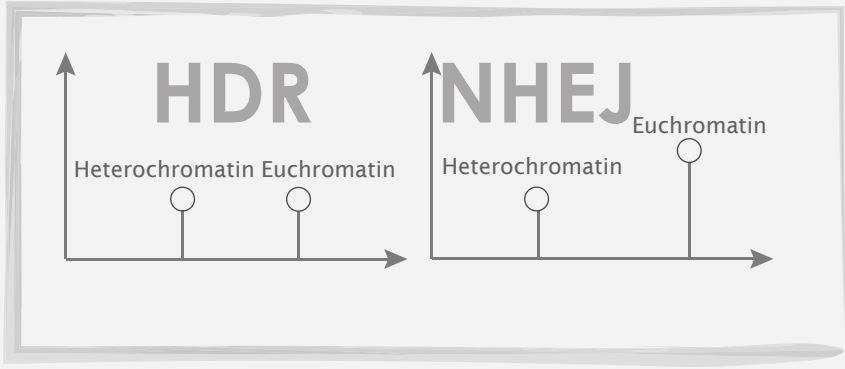
References

1. 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. 2012;109:E2579–E2586.
2. 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. 2012;337:816–821.
3. Ran F.A., Cong L., Yan W.X., Scott D.A., Gootenberg J.S., Kriz A.J., Zetsche B., Shalem O., Wu X., Makarova K.S. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520:186–191.
4. Kleinstiver B.P., Prew M.S., Tsai S.Q., Nguyen N.T., Topkar V.V., Zheng Z., Joung J.K. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. 2015;33:1293–1298.
5. Maggio I., Gonçalves M.A. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 2015;33:280–291.
6. Wright A.V., Nuñez J.K., Doudna J.A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell. 2016;164:29–44.
7. Carroll D. Staying on target with CRISPR-Cas. Nat. Biotechnol. 2013;31:807–809.
8. Fu Y., Sander J.D., Reyon D., Cascio V.M., Joung J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 2014;32:279–284.
9. Kleinstiver B.P., Pattanayak V., Prew M.S., Tsai S.Q., Nguyen N.T., Zheng Z., Joung J.K. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529:490–495.
10. Slaymaker I.M., Gao L., Zetsche B., Scott D.A., Yan W.X., Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351:84–88.
11. 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. 2013;31:833–838.
12. Ran F.A., Hsu P.D., Lin C.Y., Gootenberg J.S., Konermann S., Trevino A.E., Scott D.A., Inoue A., Matoba S., Zhang Y., Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154:1380–1389.
13. Kuscu C., Arslan S., Singh R., Thorpe J., Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat. Biotechnol. 2014;32:677–683.

14. Wu X., Scott D.A., Kriz A.J., Chiu A.C., Hsu P.D., Dadon D.B., Cheng A.W., Trevino A.E., Konermann S., Chen S. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 2014;32:670–676.
15. Polstein L.R., Perez-Pinera P., Kocak D.D., Vockley C.M., Bledsoe P., Song L., Safi A., Crawford G.E., Reddy T.E., Gersbach C.A. Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. Genome Res. 2015;25:1158–1169.
16. O'Geen H., Henry I.M., Bhakta M.S., Meckler J.F., Segal D.J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 2015;43:3389–3404.
17. Horlbeck M.A., Witkowsky L.B., Guglielmi B., Replogle J.M., Gilbert L.A., Villalta J.E., Torigoe S.E., Tjian R., Weissman J.S. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. eLife. 2016;5:e12677.
18. Chen X., Rinsma M., Janssen J.M., Liu J., Maggio I., Gonçalves M.A. Probing the impact of chromatin conformation on genome editing tools. Nucleic Acids Res. 2016;44:6482–6492.
19. Daer R.M., Cutts J.P., Brafman D.A., Haynes K.A. The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. ACS Synth. Biol. 2017;6:428–438.
20. Certo M.T., Ryu B.Y., Annis J.E., Garibov M., Jarjour J., Rawlings D.J., Scharenberg A.M. Tracking genome engineering outcome at individual DNA breakpoints. Nat. Methods. 2011;8:671–676.
21. Sternberg S.H., Redding S., Jinek M., Greene E.C., Doudna J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507:62–67.
22. Dahlman J.E., Abudayyeh O.O., Joung J., Gootenberg J.S., Zhang F., Konermann S. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. Nat. Biotechnol. 2015;33:1159–1161.
23. Kiani S., Chavez A., Tuttle M., Hall R.N., Chari R., Ter-Ovanesyan D., Qian J., Pruitt B.W., Beal J., Vora S. Cas9 gRNA engineering for genome editing, activation and repression. Nat. Methods. 2015;12:1051–1054.

Chapter 5

The **Chromatin** Structure **Governs** Gene-editing Outcomes



Josephine M. Janssen*, Xiaoyu Chen*, Jin Liu and Manuel A.F.V. Gonçalves

*Co-first Authors

Abstract

Gene editing based on homology-directed repair (HDR) depends on donor DNA and sequence-specific programmable nucleases. However, in addition to inducing HDR involving the mending of chromosomal double-stranded breaks (DSBs) with donor DNA templates, programmable nucleases also yield gene disruptions by triggering the competing non-homologous end-joining (NHEJ) pathway. Hence, it is crucial to identify parameters underlying the choice between these two DNA repair pathways in the context of HDR-based gene editing. Here, we implemented quantitative cellular systems based on epigenetically regulated isogenic target sequences and donor DNA of viral, non-viral and synthetic origins, to investigate gene editing outcomes resulting from the interaction between different chromatin conformations and donor DNA structures. We demonstrate that the chromatin topology influences gene editing endpoints by shifting the balance between HDR and NHEJ events. In particular, HDR increases in relation to NHEJ when chromosomal target sequences acquire a heterochromatic state. Moreover, albeit varying in degree, this shift in the balance between HDR- and NHEJ-induced chromosomal changes (1.5- to 6.4-fold) takes place independently of the types of episomal donor DNA. Besides establishing a direct relationship between specific gene editing outcomes and epigenetically regulated higher-order chromatin "conformers", these findings might guide the development of improved genome engineering procedures.

Introduction

Genome editing based on inducing targeted chromosomal double-stranded DNA breaks (DSBs) by programmable nucleases permits altering, in a precise manner, the genetic make-up of eukaryotic cells. [1, 2] Normally, homology-directed repair (HDR) is the DSB repair pathway that participates in the targeted addition of new genetic information. In this case, exogenous DNA templates sharing sequences identical to chromosomal acceptor sites serve as surrogate HDR substrates for repairing the underlying sequence-specific DSBs. Ultimately, this co-option of HDR yields precise genetic alterations at predefined genomic sequences.[1, 2]

Despite its patent usefulness, HDR-based gene editing is limited by the fact that, in mammalian cells, DSBs are primarily repaired through competing non-homologous end-joining (NHEJ) pathways instead of through HDR.[3, 4] Moreover, HDR is commonly restricted to the mitotic G2/S phases of the cell cycle, when allelic sister chromatid sequences become available, while NHEJ, involving simply end-to-end ligation of broken chromosomal termini, takes place throughout the various stages of the cell cycle.[3, 4] Critically, NHEJ-mediated DSB repair often leads to the incorporation of small insertions and deletions (indels) at the target site resulting in disruptive and potentially deleterious byproducts, e.g., chromosomal translocations

and/or allelic mutations. Hence, it is important to expand our knowledge about the parameters governing the choice between these two major DNA repair pathways which, together, determine the performance of HDR-based gene editing and genomic DNA stability.

Chromatin is formed in the nucleus of eukaryotic cells by a dynamic association between genomic DNA and various types of molecules, including, histones and non-histone proteins. The basic unit of chromatin, the nucleosome, consists of ~ 147 bps of double helix wrapped around an octamer of the four core histones H3, H4, H2A and H2B.[5] The transition from compact, or "closed", heterochromatin to relaxed, or "open", euchromatin is controlled through a large number of macromolecular complexes and their respective catalytic activities, which include methylation-demethylation, acetylation-deacetylation and phosphorylation-dephosphorylation.[5] Recently, our laboratory and that of others reported that NHEJ-mediated repair of single DSBs induced by programmable nucleases can be modulated by distinct chromatin structures.[6,7] As of yet, however, the role played by such 3D structures on the performance of HDR-based gene editing has not been assessed. To address this matter, here, we sought to specifically investigate whether distinct higher-order chromatin conformations control gene editing outcomes by changing the balance between HDR and NHEJ at single, site-specific, DSBs. For these experiments, we combined programmable RNA-guided nucleases (RGN) based on the type II clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 adaptive immune system from *S. pyogenes*,[8] with donor HDR substrates of viral, non-viral and synthetic origins. In particular, as donors, we tested integrase-defective lentiviral vector genomes (IDLVs),[9] conventional recombinant plasmids and chemically synthesized single-stranded oligodeoxyribonucleotides (ODNs) with both polarities. RGNs are ribonucleoproteins formed by a complex between a fixed Cas9 protein and a flexible guide RNA (gRNA). Typically, the 5'-terminal 20 nucleotides of the gRNA (spacer) are tailored to hybridize to a chromosomal target sequence located next to a protospacer adjacent motif (PAM; NGG in the case of *S. pyogenes* Cas9). The PAM sequence signals the position for the initial protein-DNA binding mediated through the PAM-interacting domain positioned on two lobes of Cas9.[10] Next, complementarity between the spacer portion of the gRNA and PAM-adjoined DNA sequences triggers DSB formation by the coordinated catalytic activation of the nuclease domains of Cas9 (i.e. HNH and RuvC).[8]

By using the aforementioned DNA, RNA and protein tools, we performed gene-editing experiments in quantitative live-cell readout systems based on human reporter cells containing chromosomal target sequences whose epigenetic statuses are controlled by small molecule drug availability.[6] We report that the proportions between gene editing endpoints resulting from the repair of site-specific DSBs by NHEJ and HDR differ in a chromatin structure-dependent manner with HDR increasing its

prominence in relation to NHEJ when target sequences transit from an euchromatic to an heterochromatic state.

Results

Gene editing experiments were carried out in HER.TLR$^{TetO.KRAB}$ and HEK.EGFP-$^{TetO.KRAB}$ cells (**Figure 1A**). These human reporter cells express the *E. coli* tetracycline trans-repressor (tTR) fused to a mammalian Krüppel-associated box domain (KRAB). The tTR and KRAB components are the DNA-binding and effector domains of the tTR-KRAB fusion product, respectively. KRAB-containing proteins belong to the largest family of zinc-finger repressors in tetrapod vertebrates whose generic role is to recruit chromatin remodeling co-repressors via their KRAB domains after binding to specific genomic sequences through their zinc-finger motifs.[15] In particular, KRAB domains interact with KRAB-associated protein 1 (KAP-1) oligomers that form a scaffold for the binding of heterochromatin protein 1 (HP-1) isoforms (i.e. HP1α, HP1β and HP1γ), histone deacetylases (i.e. HDAC1 and HDAC2), the nucleosome remodeling factor CHD3 and the SET-domain histone methyl-transferase SETDB1 that lead to the recruitment of additional HP1 molecules via tri-methylation of lysine 9 on histone H3 (H3K9me3).[16] Ultimately, these large protein-DNA assemblies create heterochromatic regions in the genome.[17] In HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cells, in the absence of Dox, the tTR-KRAB fusion protein binds to its cognate *TetO* sequences and recruits via its KRAB repressor domain the endogenous epigenetic silencing apparatus involving, amongst other chromatin remodeling factors, KAP-1 and HP-1 (**Figure 1B**). Conversely, in the presence of Dox, tTR-KRAB suffers a conformational change that releases it from the *TetO* sequences, resulting in the transition of associated sequences from a compacted heterochromatic state (H3K9me3 high; H3-Ac low) into a relaxed euchromatic state (H3-Ac high; H3K9me3 low).[6]

We reasoned that the complementary gain-of-function and loss-of-function assays offered by HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{Tet.KRAB}$ cells should be particularly suited for assessing the impact of epigenetically regulated chromatin conformations on specific gene editing endpoints. This is so owing to the fact that these live-cell systems permit the simultaneous quantification of HDR and NHEJ events at isogenic target sequences located either in euchromatin or heterochromatin depending on the presence or absence of Dox, respectively (**Figure 1B**). Indeed, in these cells, Dox availability regulates the tTR-KRAB-mediated recruitment of the aforementioned endogenous chromatin remodeling complexes to *TetO* sequences associated with each of the reporter alleles, i.e., *TLR^{TetO}* and *EGFP^{TetO}* (**Figure 1A**).

HDR-based gene editing experiments were started by transfecting HER.TLR$^{TetO.KRAB}$ cells, cultured in the absence or in the presence of Dox, with expression plasmids encoding the RGN complex Cas9:gTLR.1 (**Supplementary Figure S1**). The target site

of Cas9:gTLR.1 is located upstream of a nonsense mutation within the *TLR*^{TetO} construct (**Figure 1A** and **Supplementary Figure S2**) and is flanked by sequences "homologous" to those present in the *EGFP*-repairing donor template *EGFPtrunc*.[12] This HDR substrate was delivered by transducing HER.TLR$^{TetO.KRAB}$ cells with different amounts of the integrase-defective lentiviral vector IDLV^{d} together with constructs expressing the RGN complex Cas9:gTLR.1 (**Figure 1B**). Negative controls were provided by HER.TLR$^{TetO.KRAB}$ cells that were neither transfected with expression plasmids nor transduced with IDLV^{d} particles (Mock) and by HER.TLR$^{TetO.KRAB}$ cells that were exposed to an irrelevant, non-targeting, gRNA (gNT) together with Cas9 and IDLV^{d}. After the action of the RGN complexes had taken place, all HER.TLR$^{TetO.KRAB}$ cultures were incubated in the presence of Dox for allowing transgene expression and quantification of HDR and NHEJ events by EGFP- and mCherry-directed flow cytometry respectively (**Figure 1A** and **Supplementary Figure S1**).

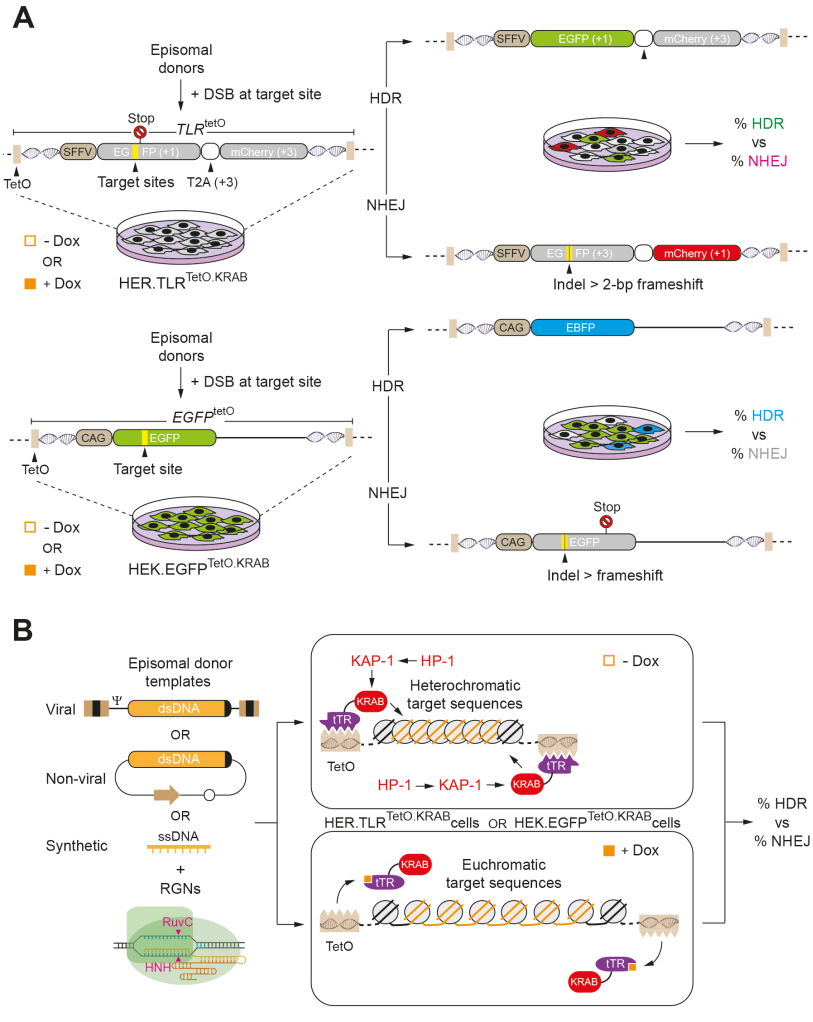


Figure 1. Experimental systems for tracking gene editing outcomes at isogenic target sequences with alternative higher-order epigenetic states. (**A**) *Modus operandi* of the cellular systems for tracking gene-editing endpoints at heterochromatin versus euchromatin. Upper panel, the *TetO*-flanked *TLR*^{TetO} construct in tTR-KRAB-expressing HER.TLR$^{TetO.KRAB}$ cells has an *EGFP* ORF interrupted by heterologous sequences and a stop codon located upstream of a *T2A* sequence and an out-of-frame *mCherry* reporter. HDR is scored by measuring EGFP+ cells resulting from the repair of site-specific DSBs by HR events between episomal donor templates (*EGFPtrunc*) and heterochromatic (-Dox) or euchromatic (+Dox) chromosomal DNA. This genetic exchange results in the substitution of the heterologous and stop codon DNA by an in-frame EGFP sequence. Concomitantly, NHEJ is scored by measuring mCherry+ cells resulting from indels placing the mCherry in-frame. Lower pane, the *TetO*-flanked *EGFP* construct (*EGFP*^{TetO}) in tTR-KRAB-expressing HEK.EGFP$^{TetO.KRAB}$ cells is functional. HDR is tracked by measuring the frequencies of blue light-emitting cells resulting from the conversion of the EGFP fluorochrome to that of EBFP. Simultaneously, NHEJ is scored by measuring EGFP- cells resulting from indels placing the *EGFP* sequence out-of-frame. (**B**) Generic experimental designs. The reporter HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cells, cultured in the absence or in the presence of Dox, are exposed to RGNs together with different donor DNA templates. Without Dox, tTR-KRAB binds to *TetO* and induces heterochromatin formation through the recruitment of, amongst others, KAP-1 and HP-1. With Dox, tTR-KRAB set free *TetO* leading the target sequences to acquire an euchromatic state. After the completion of the gene editing processes, Dox is added to the different cultures in order to determine the frequencies of HDR and NHEJ events at heterochromatic versus euchromatic target sequences by dual-color flow cytometry.

The results obtained from this experiment revealed that the frequencies of DSB-triggered NHEJ at euchromatic target sequences (+Dox) were substantially higher than those measured at their heterochromatic (-Dox) counterparts as assessed by mCherry-directed flow cytometry (**Figures 2A** and **2**B). This outcome is in agreement with that of our previous study involving the exclusive delivery of RGNs into HER.TLR$^{TetO.KRAB}$ cells.[6] In particular, RGN-induced DSBs are preferentially formed at euchromatin over heterochromatin,[6] which, in turn, correlates with the preferential binding of RGNs harboring catalytically inert ("dead") Cas9 proteins to euchromatic over heterochromatic regions across the genome.[18-20] Interestingly, despite of the initial higher accessibility of gene editing tools to euchromatic over heterochromatic genomic DNA, there were no corresponding increases in HDR levels in the former, Dox-treated, cells (**Figures 2A** and **2B**). As a result, the ratios between NHEJ and HDR events at compact heterochromatin were substantially lower than those measured at relaxed euchromatin (4.6- to 5.3-fold), regardless of the amounts of exogenous HDR templates available for recombination (**Figure 2C**, top graph). This outcome translated in a relative increase in HDR (+) and a decrease in NHEJ (-) at heterochromatin (**Figure 2C**, bottom graph). The use of the alternative RGN complex Cas9:gTLR.2 (**Figure 3A** and **Supplementary Figure S3**) and a different transfection protocol (**Figures 3B** and **3C**), led to similar NHEJ/HDR ratios and variations in the frequencies of HDR and NHEJ (compare **Figure 3D** with **Figure 2C**, respectively).

Next, we sought to determine RGN-induced gene editing endpoints at isogenic target sequences with distinct higher-order chromatin conformations after delivering donor DNA in the context of covalently closed double-stranded plasmids. In these

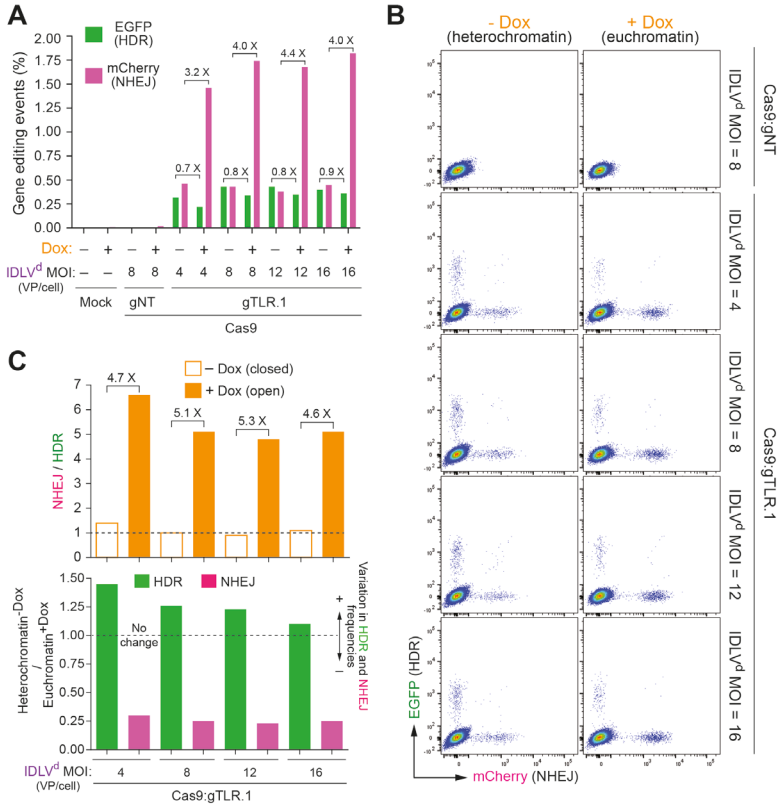


Figure 2. Gene editing endpoints at euchromatin versus heterochromatin after IDLV donor DNA delivery. (**A**) Dual-color flow cytometric quantification of HDR and NHEJ events in HER.TLR$^{TetO.KRAB}$ cells. HER.TLR$^{TetO.KRAB}$ cells were exposed to Cas9:gTLR.1 together with the indicated multiplicities of infection (MOI) of IDLV^{d}. Negative controls consisted of mock-treated cultures and of cultures exposed to a non-targeting gRNA (gNT), Cas9 and IDLV^{d} at an MOI of 8 vector particles per cell (VP/cell). The various experimental conditions were tested in HER.TLR$^{TetO.KRAB}$ reporter cells incubated in the absence (-) or in the presence (+) of doxycycline (Dox). The frequencies of HDR and NHEJ events in the various target cell populations were determined by measuring EGFP+ and mCherry+ cells, respectively. (**B**) Dot plots corresponding to HER.TLR$^{TetO.KRAB}$ cells transduced with different doses of IDLV^{d} particles and subjected to the indicated Dox regimens. (**C**) Relative participation of HDR and NHEJ pathways during IDLV-mediated repair of DSBs occurring at heterochromatin versus euchromatin. In the top graph, data of panel **A** are presented as the ratios between the frequencies of NHEJ and HDR in HER.TLR$^{TetO.KRAB}$ cells not treated and treated with Dox (top graph). In the bottom graph, data of panel **A** are depicted as the variation in the proportion of HDR and NHEJ events at heterochromatin versus euchromatin.

HDR during heterochromatic DSB repair

experiments, we deployed the lentiviral DNA construct Plasmid^{d},[12] which had been utilized for assembling IDLV^{d} particles. Again, these gene editing experiments involved the use of two different transfection protocols for introducing donor Plasmid^{d} mixed with constructs expressing either Cas9:gTLR.1 or Cas9:gTLR.2 complexes into HER.TLR$^{TetO.KRAB}$ cells treated or not treated with Dox (**Figures 3E-G**). The resulting gene editing outcomes (**Figures 3E-H**) were similar to those obtained after

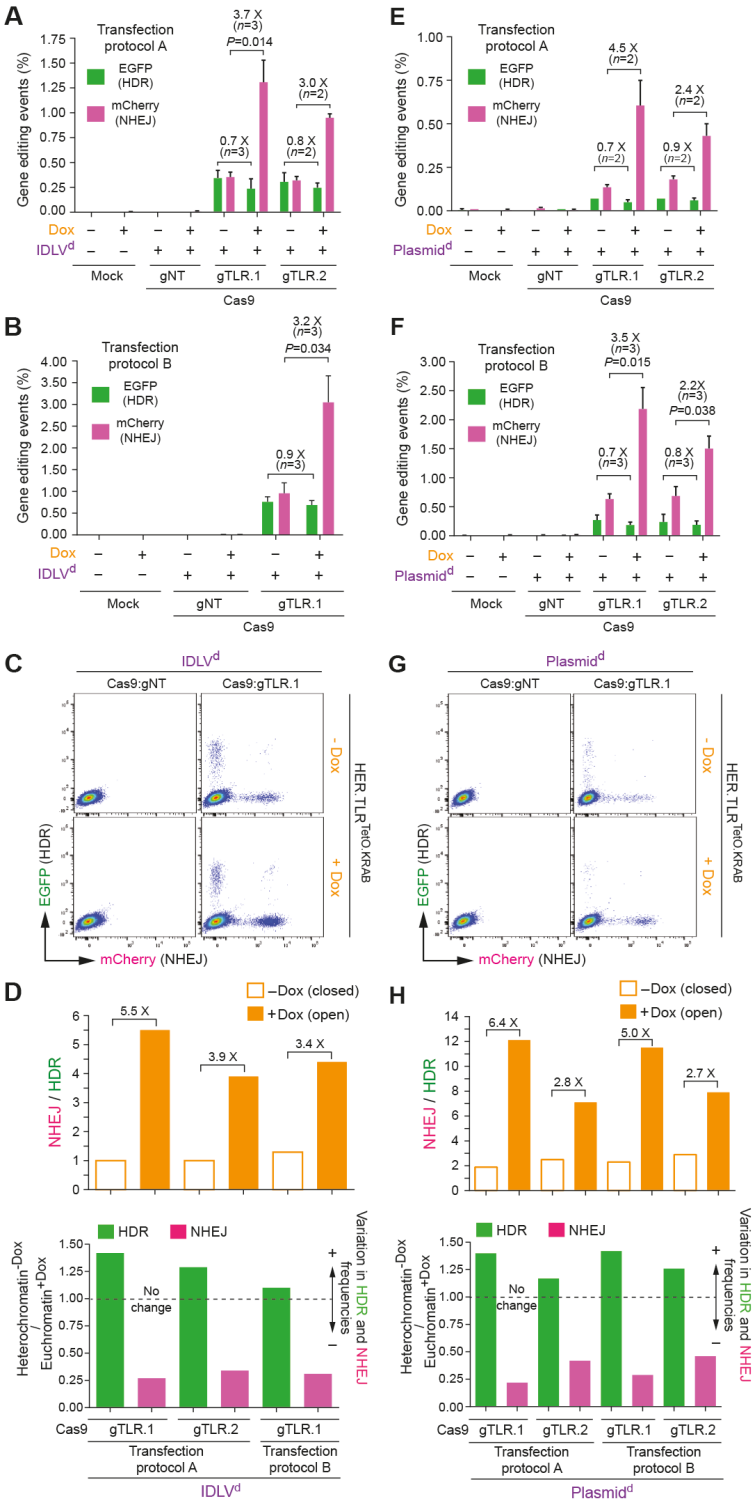

A
Transfection protocol A
EGFP (HDR)
mCherry (NHEJ)
Gene editing events (%)
3.7 X (n=3) P=0.014
0.7 X (n=3)
0.8 X (n=2)
3.0 X (n=2)
Dox
IDLV^d
Mock
gNT
gTLR.1
gTLR.2
Cas9
E
Transfection protocol A
4.5 X (n=2)
0.7 X (n=2)
0.9 X (n=2)
2.4 X (n=2)
Plasmid^d
B
Transfection protocol B
3.2 X (n=3) P=0.034
0.9 X (n=3)
F
Transfection protocol B
3.5 X (n=3) P=0.015
0.7 X (n=3)
2.2X (n=3) P=0.038
0.8 X (n=3)
C
IDLV^d
Cas9:gNT
Cas9:gTLR.1
- Dox
+ Dox
HER.TLR^TetO KRAB
EGFP (HDR)
mCherry (NHEJ)
G
Plasmid^d
D
–Dox (closed)
+Dox (open)
NHEJ / HDR
5.5 X
3.9 X
3.4 X
HDR
NHEJ
Heterochromatin -Dox / Euchromatin +Dox
No change
Variation in HDR and NHEJ frequencies
Transfection protocol A
Transfection protocol B
IDLV^d
H
6.4 X
2.8 X
5.0 X
2.7 X
Plasmid^d

Figure 3. Comparing gene editing outcomes at euchromatin versus heterochromatin after viral and plasmid vector delivery of donor DNA. (**A** and **B**) IDLV^{d}-based gene editing. Dual-color flow cytometric measurements of HDR and NHEJ frequencies in HER.TLR$^{TetO.KRAB}$ cells subjected to the indicated experimental conditions and treated (+) or not treated (-) with Dox. Two different transfection protocols (**A** and **B**) were used to introduce the DNA constructs into target cells. IDLV^{d} particles were applied at an MOI of 8 VP/cell. (**C**) Representative dot plots corresponding to HER.TLR$^{TetO.KRAB}$ cells exposed to IDLV^{d} together with Cas9:gNT or Cas9:gTLR.1 complexes. (**D**) Comparative engagement of HDR and NHEJ pathways during IDLV-mediated repair of DSBs made at heterochromatin versus euchromatin. Top graph, data of panels **A** and **B** presented as the ratios between the rates of NHEJ and HDR in HER.TLR$^{TetO.KRAB}$ cells either incubated or not incubated with Dox. Bottom graph, data of panels **A** and **B** depicted as the variation in the fraction of HDR and NHEJ events at heterochromatin versus euchromatin. (**E** and **F**) Plasmid^{d}-based gene editing. Dual-color flow cytometric quantification of HDR and NHEJ frequencies in HER.TLR$^{TetO.KRAB}$ cells. HER.TLR$^{TetO.KRAB}$ cells incubated (+) or not incubated (-) with Dox, were either mock-transfected or were transfected with Plasmid^{d} mixed with constructs encoding the indicated RGN complexes. Two different transfection protocols (**A** and **B**) were used to deliver the DNA constructs into target cells. (**G**) Representative dot plots corresponding to HER.TLR$^{TetO.KRAB}$ cells transfected with Plasmid^{d} mixed with expression constructs coding for Cas9:gNT or Cas9:TLR.1 complexes. (**H**) Relative engagement of HDR and NHEJ pathways during plasmid-mediated repair of DSBs created at heterochromatin versus euchromatin. Top graph, results of panels **E** and **F** depicted as ratios between the frequencies of NHEJ and HDR in HER.TLR$^{TetO.KRAB}$ cells exposed or not exposed to Dox. Bottom graph, data of panels **E** and **F** shown as the variation in HDR and NHEJ events at heterochromatin versus euchromatin. Bars in graphs **A**, **B**, **E** and **F** correspond to mean ± s.d. of the indicated number (*n*) of independent experiments (biological replicates done in different days).

IDLV^{d} transduction of HER.TLR$^{TetO.KRAB}$ cells (**Figure 2** and **Figures 3A-D**). In particular, in comparison with euchromatin, at heterochromatin, the balance between NHEJ and HDR shifts towards the latter pathway causing target cell populations to acquire a more even distribution between HDR- and NHEJ-derived genetic modifications (**Figure 3H**).

To serve as additional controls, gene editing experiments were also performed in tTR-KRAB-expressing HER.TLR^{KRAB} cells whose target sequences are not under conditional KRAB-mediated epigenetic regulation due to their lack of *TetO cis*-acting elements necessary for tTR-KRAB binding (**Figure 4A**). Importantly, regardless of the Dox regiment, neither the HDR levels nor the NHEJ levels changed in HER.TLR^{KRAB} cells, independently of whether the donor DNA was introduced into target cell nuclei in the context of linear IDLV^{d} genomes (**Figures 4B** and **4C**) or covalently closed Plasmid^{d} molecules (**Figures 4D** and **4E**). Hence, in contrast with gene editing experiments in HER.TLR$^{TetO.KRAB}$ cells, in control HER.TLR^{KRAB} cells, there were no substantial Dox-dependent variations in the proportions between HDR and NHEJ events for both types of donor DNA templates used (**Figure 4F**).

Next, we performed gene-editing experiments in HEK.EGFP$^{TetO.KRAB}$ cells. In this independent experimental system, HDR can be promptly tracked by measuring cells in which the EGFP fluorochrome is converted into that of EBFP, while NHEJ can be monitored through quantifying cells with indel-derived EGFP knockouts (**Figures**

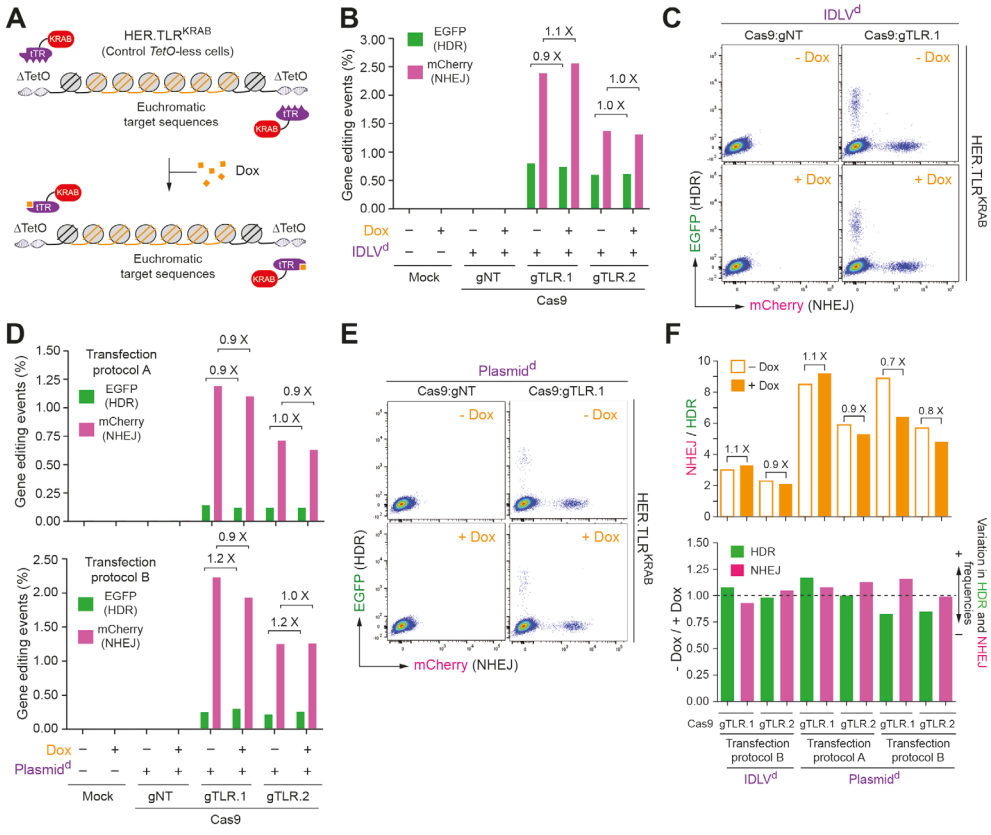


Figure 4. Gene editing endpoints in control HER.TLR^{KRAB} cells exposed or not exposed to Dox. (**A**) Schematics of target DNA in HER.TLR^{KRAB} cells. The tTR-KRAB-expressing HER.TLR^{KRAB} cells have a Dox-insensitive *TLR* construct due to its lack of *cis*-acting *TetO* elements. (**B**) Dual-color flow cytometric quantification of HDR and NHEJ events in HER.TLR^{KRAB} cells. HER.TLR^{KRAB} cells, treated (+) or not treated (-) with Dox, were exposed to the indicated experimental conditions. IDLV^{d} particles were applied at an MOI of 8 VP/cell. (**C**) Representative dot plots corresponding to HER.TLR^{KRAB} cells exposed to IDLV^{d} particles together with Cas9:gNT or Cas9:TLR.1 complexes. (**D**) Dual-color flow cytometric quantification of HDR and NHEJ frequencies in HER.TLR^{KRAB} cells. HER.TLR^{KRAB} cells, incubated (+) or not incubated (-) with Dox, were mock transfected or were transfected with plasmid mixed with constructs encoding the indicated RGN complexes. Two different transfection protocols (A and B) were used to deliver the DNA constructs into target cells. (**E**) Dot plots corresponding to HER.TLR^{KRAB} cells transfected with Plasmid^{d} mixed with expression constructs coding for Cas9:gNT or Cas9:TLR.1 complexes. (**F**) Comparative engagement of HDR and NHEJ pathways at site-specific DSBs created at heterochromatin versus euchromatin. Top graph, data of panels **B** and **D** presented as ratios between the rates of NHEJ and HDR in HER.TLR^{KRAB} cells not incubated or incubated in the presence of Dox. Bottom graph, data of panels **B** and **D** shown as the variation in the fraction of HDR and NHEJ events at heterochromatin versus euchromatin.

5A and **5B**). In these experiments, HEK.EGFP$^{TetO.KRAB}$ cells, cultured in the absence or in the presence of Dox, were transfected with plasmid pTHG.Donor together with constructs encoding the Cas9:gEGFP complex targeting the *EGFP* fluorochrome coding sequence (**Figure 5A**). These data were in agreement with those obtained in HER.TLR$^{TetO.KRAB}$ cells (**Figures 2** and **3**) in that, notwithstanding the higher frequencies of site-specific DSBs at euchromatin over those measured at heterochromatin, HDR levels were comparable at both chromatin states (**Figure 5C**). As a result, the NHEJ/HDR ratios at heterochromatin were consistently lower than those measured at euchromatin (**Figure 5D**, top graph), yielding a relative increase in HDR and a simultaneous decrease in NHEJ at the former chromatin state (**Figure 5D**, bottom graph).

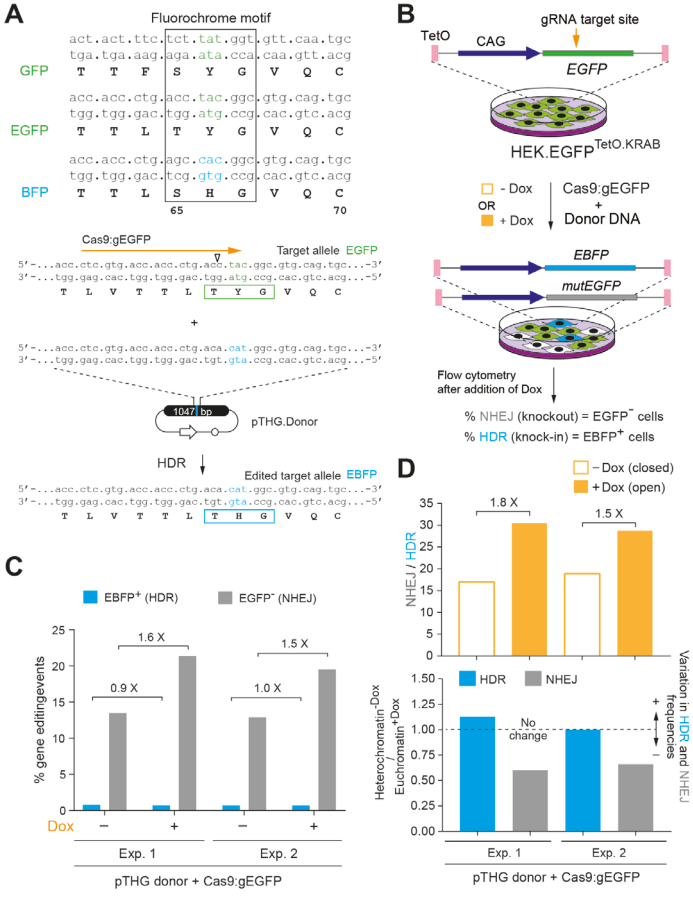


Figure 5. Gene editing outcomes at euchromatin versus heterochromatin after plasmid donor delivery into HEK.EGFP$^{TetO.KRAB}$ cells. (**A**) Gene editing assay based on EGFP-to-EBFP fluorochrome conversion. Top panel, nucleic acid and amino acid sequences corresponding to the fluorochromes of GFP, EGFP and BFP (*boxed*). Bottom panel, nucleotide and amino acid sequences of the reporter target allele before and after its editing through the delivery of pTHG.Donor and expression constructs encoding the RGN complex Cas9:gRNA^{EGFP}. *Horizontal orange arrow*, target site of Cas9:gRNA^{EGFP}; *vertical open arrowhead*, position of the

DSB induced by Cas9:gRNA^{EGFP}. (**B**) Schematics of the experimental design applied to HEK. EGFP$^{TetO.KRAB}$ cells. (**C**) Flow cytometric quantification of HDR and NHEJ frequencies. HEK. EGFP$^{TetO.KRAB}$ cells, incubated (+) or not incubated (-) with Dox, were exposed to pTHG.Donor and gRNA^{EGFP}-containing RGNs. The frequencies of HDR and NHEJ events in the transfected cell populations were determined by measuring EBFP+ and EGFP- cells, respectively. A minimum of forty thousand events, each corresponding to a single viable cell, were acquired per sample. (**D**) Relative participation of HDR and NHEJ pathways during plasmid-mediated repair of DSBs made at heterochromatin versus euchromatin. Top graph, data of panel **C** presented as the ratios between the frequencies of NHEJ and HDR in HEK.EGFP$^{TetO.KRAB}$ cells treated and not treated with Dox. Bottom graph, data of panel **C** depicted as the variation in the proportion of HDR and NHEJ events at heterochromatin versus euchromatin.

Finally, to complement the previous experiments testing linear and covalently closed double-stranded donors in the form of IDLV genomes and recombinant plasmids, respectively, we sought to assess ODN-based gene editing at euchromatin versus heterochromatin. For these experiments, we selected a single-stranded ODN pair corresponding to the sense and antisense polarities of the target polynucleotide chains of Cas9:gEGFP (i.e. ODN.s and ODN.as, respectively). Previous research has demonstrated that RGNs can display a long residence time on target DNA (~ 6 h) and that, after DNA cutting, the strand upstream of the PAM (non-target strand) is released from the Cas9-gRNA-DNA ternary complex forming a 3'-ended DNA flap (**Figure 6A**).[21] This insight permitted the design of optimized single-stranded ODN donors which are complementary to the released strand. Indeed, when compared to double-stranded and single-stranded ODNs that cannot anneal to RGN-generated flaps, ODNs complementary to the released strand induced ~ 4- and ~ 2-fold higher frequencies of HDR in human cells, respectively.[21] Results from an initial experiment in HEK.EGFP$^{TetO.KRAB}$ cells exposed to Cas9:gEGFP together with ODN.s or with ODN.as were consistent with the aforementioned data in that the ODN.as yielded ~2-fold higher frequencies of HDR than the ODN.s (**Figure 6B**). Interestingly, expanding these ODN transfection experiments to HEK.EGFP$^{TetO.KRAB}$ cells treated or not treated with Dox revealed that, at both euchromatin and heterochromatin, the flap-hybridizing donor ODN.as consistently yielded a more even distribution between HDR and NHEJ events when compared to its ODN.s counterpart (**Figures 6C** and **6D**). These data suggest that base pairing assists in the engagement of flap-annealing ODNs with the RGN-cleaved target site dampening the contribution of the NHEJ pathway to the repair of the underlying site-specific DSBs. Importantly, when comparing ODN-based gene editing endpoints at euchromatin versus heterochomatin, these and follow-up ODN.as dose-response experiments were in agreement with the previous experiments using IDLV and plasmid donor DNA (**Figure 6E**). In particular, the frequencies of HDR and NHEJ were more comparable at heterochromatin than at euchromatin independently of ODN.as concentrations (**Figures 6F**, top graph). As a result, when target DNA sequences transit from an euchromatic to an heterochromatic state, there is a shift towards an increase in the preponderance of HDR over NHEJ (**Figrue 6F**, bottom graph).

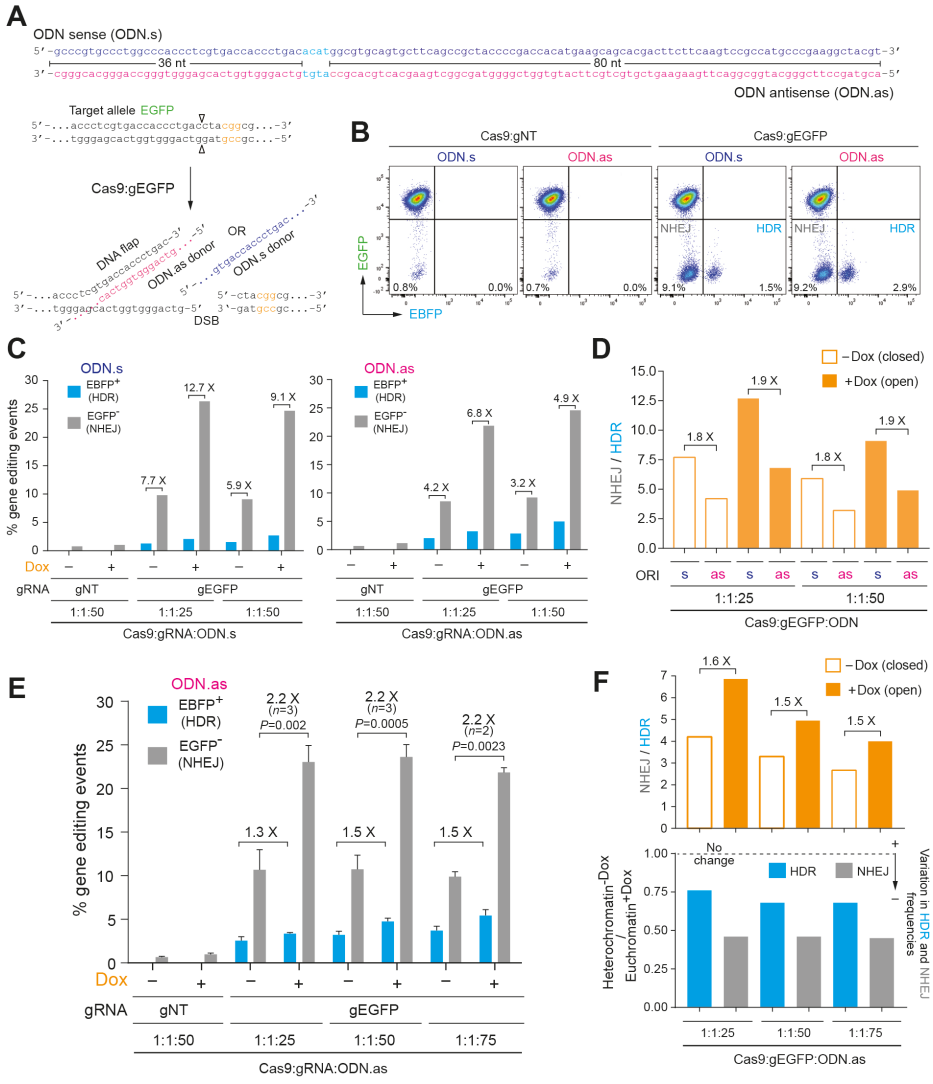


Figure 6. Gene editing endpoints at euchromatin versus heterochromatin after ODN donor delivery in HEK.EGFP$^{TetO.KRAB}$ cells. (**A**) Schematics of ODN design and target site before and after RGN engagement. The RGN complex Cas9:gEGFP is presumed to generate a 3'-ended DNA flap complementary and non-complementary to ODN.as and ODN.s, respectively. HDR-based gene editing with ODN.s and ODN.as donors should result in EGFP-to-EBFP conversion. *Open arrowheads*, position of the DSB induced by Cas9:gEGFP. *Orange triplet*, PAM. (**B**) Probing HDR-based gene editing with sense and antisense ODNs. HEK.EGFP$^{TetO.KRAB}$ cells were transfected with ODN.s or with ODN.as each mixed with expression plasmids coding for either non-cutting Cas9:gNT or cutting Cas9:gEGFP complexes. HDR and NHEJ quantification in HEK.EGFP$^{TetO.KRAB}$ cells was assessed by EBFP- and EGFP-directed flow cytometry, respectively. (**C**) Testing the impact of chromatin structure on HDR-based gene editing with sense and antisense ODNs. HEK.EGFP$^{TetO.KRAB}$ cells, incubated (+) or not incubated (-) with Dox, were exposed to the indicated experimental conditions. The frequencies of HDR and NHEJ were assessed by dual-color flow cytometry. (**D**) Relative participation of HDR and NHEJ pathways during the repair of euchromatic versus heterochromatic DSBs with ODNs with different polarities. Data of panel **C** displayed as the ratios between the frequencies of

NHEJ and HDR in HEK.EGFP$^{TetO.KRAB}$ cells treated and not treated with Dox. (**E**) ODN-based gene editing. Dual-color flow cytometric quantification of HDR and NHEJ frequencies in HEK.EGFP$^{TetO.KRAB}$ cells. HER.TLR$^{TetO.KRAB}$ cells incubated (+) or not incubated (-) with Dox, were exposed to the indicated experimental conditions. Bars correspond to mean ± s.d. of the indicated number (*n*) of independent experiments (biological replicates done in different days). (**F**) Relative participation of HDR and NHEJ pathways during ODN-mediated repair of DSBs taking place at heterochromatin versus euchromatin. Top graph, results of panel **E** shown as the ratios between the frequencies of NHEJ and HDR in HEK.EGFP$^{TetO.KRAB}$ cells exposed or not exposed to Dox. Bottom graph, data of panel **E** presented as the variation in HDR and NHEJ events at heterochromatin versus euchromatin.

Taken our data together, we conclude that site-specific DSBs generated within euchromatin are mostly repaired through mutagenic NHEJ in detriment of error-free HDR. However, if the site-specific DSBs are made within heterochromatin instead, there is a more balanced participation of both cellular machineries in the repair of site-specific DNA lesions (**Figure 7**). Albeit varying in degree, this chromatin structure-dependent shift in the relationship between NHEJ and HDR takes place regardless of whether the donor DNA is presented in the context of IDLV genomes, recombinant plasmids or single-stranded ODNs, which together, makeup the most commonly used sources of exogenous genetic information.

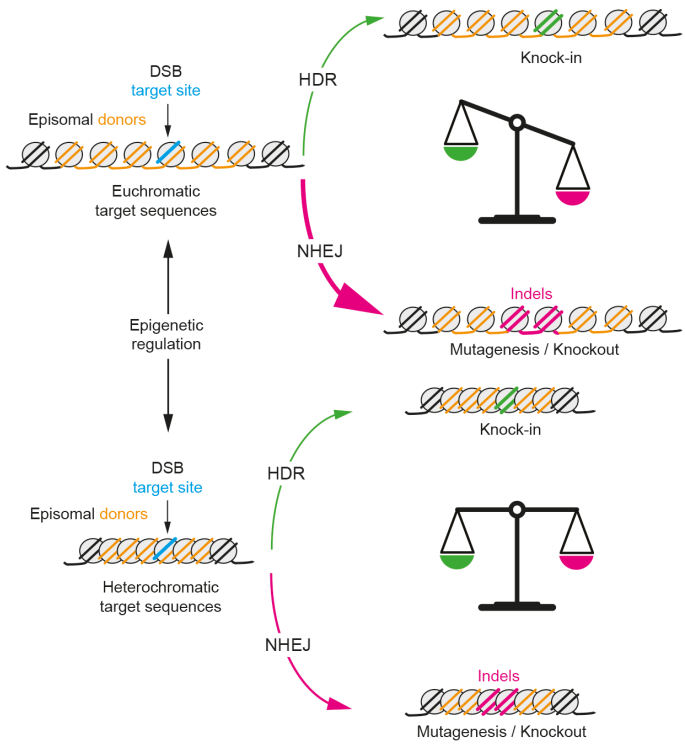


Figure 7. Summarizing illustration on the role of the chromatin structure on gene editing outcomes. The thickness of the curved arrows represents the relative contribution of homology-directed repair (HDR) and non-homologous end-joining (NHEJ) to gene editing endpoints at euchromatin versus heterochromatin.

Discussion

HDR-based genome editing is crucial for numerous research applications, including modelling, screening or correcting genotypes underlying human disorders in stem and/or progenitor cells. Crucially, accurate HDR takes place much less frequently than mutagenic NHEJ.[3,4] Thus, identifying the biological parameters governing this strong DNA repair bias has both scientific and practical relevance. In this study, we have investigated the outcome of the interaction between the molecular tools necessary for HDR-based gene editing and the chromatin structure of target sequences. In particular, we assessed RGN-induced gene editing endpoints established after the engagement of donors of viral, non-viral and synthetic origins with isogenic target sequences located either in euchromatin or heterochromatin. We found that the relative proportions of gene editing endpoints resulting from mutagenic NHEJ and precise HDR events depend to a significant degree on the higher-order chromatin conformation of target sequences with a shift occurring towards HDR events at heterochromatin (**Figure 7**). This bias can vary in its extent, such as when using ssODNs with different polarities (~2-fold; **Figures 6C** and **6D**), but takes place independently of the type of episomal donor DNA utilized.

These findings suggest that HDR-based gene editing can be impacted by the epigenomic landscape of specific cell types as well as by the dynamic and epigenetically regulated chromatin changes underlying organismal development and cellular differentiation stages. Indeed, our experimental results support the hypothesis that the chromatin environment contributes to the well-known differential susceptibility of genomic sequences to gene editing interventions. Hence, the chromatin context of the target sequence should be taken into account whenever considering applying HDR-based gene editing procedures.

There is a paucity of knowledge about the repair mechanisms of DSBs located within different chromatin contexts in mammalian cells. In recent years, however, the classical view that heterochromatin simply poses a barrier to the DNA damage response (DDR) is changing into one in which heterochromatin and heterochromatin-associated proteins are active participants in it.[22] For instance, SENP7 interacts with KAP-1 via HP1α resulting in the deSUMOylation of KAP-1.[23] The removal of this post-translational modification from KAP-1 promotes the transient release of the co-repressors CHD3 and SETDB1 from chromatin, which in turn, creates a cellular milieu favorable for HDR-mediated DSB repair.[23] A similar milieu is conferred by the MRN-dependent recruitment of the histone acetyltransferase Trrap-Tip60 to

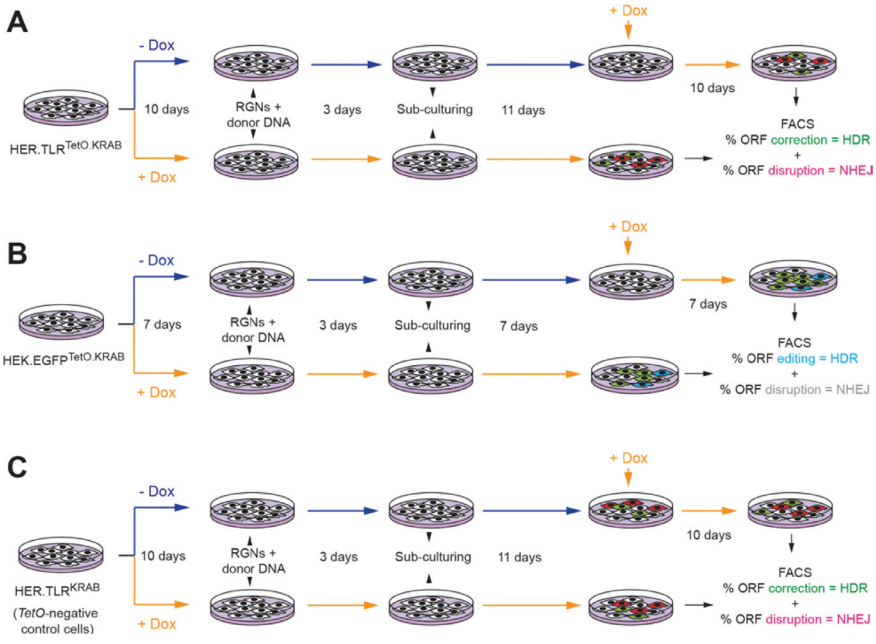
heterochromatic DSBs.[24] It has also been shown that HP1α is transiently mobilized to both euchromatic and heterochromatic DSBs via an interaction with p150CAF-1, resulting in its higher accumulation at the latter lesions.[25] Interestingly, in HP1α knockdown cells, in contrast to the buildup of the NHEJ factor XRCC4 at laser-induced DNA lesions, there is a markedly reduction of the HDR factors RAD51 and BRCA1 at these lesions.[25] Subsequent experiments, based on exposing cells to the restriction enzyme AsiSI, provided additional support for the participation of heterochromatin-resident HP1 proteins in associating BRCA1 with DSBs and facilitating HDR.[26]

Recent experiments are also starting to shed light on the relationship between different cell cycle stages and DNA repair pathways at heterochromatic domains in mammalian cells. Study showed that DSBs created within pericentric heterochromatin during G1 remain stationary and are repaired through NHEJ, whilst in S or G2, these DSBs relocate to the periphery of the heterochromatic domain and, once there, become substrates for RAD51/BRCA2-dependent HDR .[27] This heterochromatic DSB migration to euchromatic regions might favor the finalization of proper HDR with sister chromatid or homologous chromosome sequences in detriment of ectopic HDR with repetitive DNA, common in heterochromatic regions. Remarkably, DSBs located within centromeric heterochromatin, recruit not only the NHEJ marker protein Ku80 but also the HDR factors RPA and RAD51 throughout the cell cycle with an enhancement observed during G2.[27]

Collectively, these data provide compelling evidence for an active role of HDR during heterochromatic DSB repair involving an intricate interplay between histone marks (e.g. H3K9me3), chromatin remodeling factors (e.g. HP1 isoforms, CHD3, Trrap-Tip60 and KAP-1) and DNA repair proteins (e.g. BRCA1, RPA and RAD51). It is worth mentioning, however, that for the most part, these experiments have relied on generating supra-physiological amounts of different types of DSBs throughout the genome either by ionizing radiation, laser micro-irradiation or restriction enzyme exposure. Moreover, the relative proportions between HDR and NHEJ events at sequences with distinct chromatin states in individual test cell populations were not investigated. Finally, although certain DDR processes seem to be specific for repairing heterochromatic DSBs, e.g., ATM-mediated phosphorylation of KAP-1, [28] some others appear to lack this specificity, e.g., p150CAF-1-mediated recruitment of HP1α to DSBs.[25] It should thus be very instructive investigating which DDR components and mechanisms are specific to heterochromatin, euchromatin or shared by both compartments.

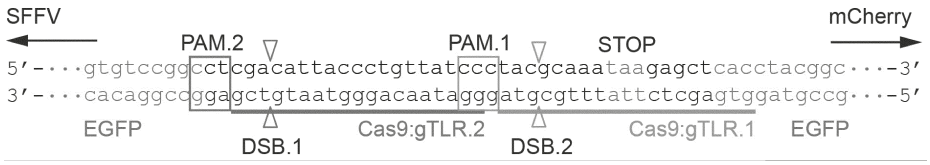
Concluding, in the present study, we have implemented cellular assays based on epigenetically regulated genetic reporters, donor DNA templates and RGNs for the simultaneous quantification of HDR and NHEJ events at single target sequences

subjected to distinct chromatin conformations. The resulting data expand the aforementioned findings by providing direct experimental evidence for a role of the higher-order chromatin structure on the differential regulation of the two major DNA repair pathways in mammalian cells. The recruitment of DDR factors and DNA recombination substrates into a well-defined genetic and epigenetic environment offered by these live-cell tracking systems should aid detailed investigations into the mechanisms of DDR under different chromatin contexts as well as their interplay with other cellular mechanisms and DNA metabolic processes such as replication. Finally, as illustrated in the current study through experiments testing viral, non-viral and synthetic donors, this epigenetically-regulated experimental systems should also serve for assessing in cellula the impact of chromatin on novel gene editing protocols involving, amongst others, donor DNA substrates from different origins or with different structures and compositions, NHEJ-inhibiting reagents,[29, 30] and unexploited programmable nuclease systems.[31, 32]



Supplementary Figure S1. Schematic representation of the experimental settings used in the current study. The tTR-KRAB-expressing cells HER.TLR$^{TetO.KRAB}$ (**A**) and HEK.EGFP$^{TetO.KRAB}$ (**B**) contain the Dox-regulated *TLR^{TetO}* [9] and *EGFP^{TetO}* [33] constructs, respectively. These reporter cells, containing target sequences in a heterochromatic (-Dox) or euchromatic (+Dox) state, are transiently transfected with different combinations of gene editing tools consisting of RGNs and donor DNA templates. After the generation of site-specific DSBs and the ensuing modification of target DNA sequences in cells subjected to both experimental settings (i.e. –Dox and +Dox), target gene expression is activated to quantifying by flow cytometry

the frequencies of gene editing events resulting from the engagement of HDR and NHEJ pathways. The tTR-KRAB-expressing HER.TLR^{KRAB} reporter cells (**C**) have the Dox-insensitive *TLR* construct [11] and were used as an isogenic control cellular system.



Supplementary Figure S2. Target sites of RGN complexes in the *TLR* construct. The target sequences for the RGN complexes Cas9:gTLR.1 and Cas9:gTLR.2 are indicated by horizontal lines linked to *open boxes* (PAM elements). The positions of the DSBs generated by each RGN are marked (*vertical open arrowheads*). STOP, nonsense codon located within the *TLR* ORF.

Supplementary Table S1. Oligonucleotide pairs to generate the gRNA expression constructs expressing gTLR.1, gTLR.2, gNT and gEGFP

Plasmids	Oligonucleotide pairs (5′- 3′)
Z42_pgTLR.1	5′-ACCGGTGAGCTCTTATTTGCGTA-3′ 5′-AAACTACGCAAATAAGAGCTCAC-3′
Z44_pTLR.2	5′-ACCGGGATAACAGGGTAATGTCG-3 5′-AAACCGACATTACCCTGTTATCC-3′
AM51_pgNT	5′-ACCGGTGAGCTCTTATTTGCGTAGCTAGCTGAC-3 5′-AAACGTCAGCTAGCTACGCAAATAAGAGCTCAC-3′
AX03_pgEGFP	5′-ACCGCTCGTGACCACCCTGACCTA-3′ 5′-AAACTAGGTCAGGGTGGTCACGAG-3′

Supplementary Table S2. Experimental scheme corresponding to **Figure 2** (**Protocol A**)

DONOR: IDLV^{d}	3.25 ×10^{5} HER.TLR$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)			
	PEI (1mg/ml) 5.8 µl per well; Ratio DNA / PEI equivalents = 6			
Reagents	Cas9	gNT (Ctrl)	gTLR.1	Total (ng)
Construct length (bp)	9551	3056	3046	
DNA per well (ng)	1327	423		1750
	1327		423	1750
	1327			1750

Note 1: One day after transfecting plasmids expressing Cas9 and gTLR.1, IDLV^{d} particles were added at an MOI of 4, 8, 12 and 16 VP/cell; **Note 2**: One day after transfecting plasmids expressing Cas9 and gNT, IDLV^{d} particles were added at an MOIs of 8 VP/cell.

Supplementary Table S3. Experimental scheme corresponding to **Figure 3A** (**Protocol A**)

DONOR: IDLV^{d}	3.25 ×10^{5} HER.TLR$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)				
	PEI (1mg/ml) 5.8 µl per well; Ratio DNA / PEI equivalents = 6				
Reagents	Cas9	gNT (Ctrl)	gTLR.1	gTLR.2	Total (ng)
Construct length (bp)	9551	3056	3046	3046	
DNA per well (ng)	1327	423			1750
	1327		423		1750
	1327			423	1750

Note: One day after transfecting the indicated plasmids, IDLV^{d} particles were added at an MOI of 8 VP/cell.

Supplementary Table S4. Experimental scheme corresponding to **Figure 3B** (**Protocol B**)

DONOR: IDLV^{d}	3.25 ×10^{5} HER.TLR$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)			
	PEI (1mg/ml) 9.6 µl per well; Ratio DNA / PEI equivalents = 10			
Reagents	Cas9	gNT (Ctrl)	gTLR.1	Total (ng)
Construct length (bp)	9551	3056	3046	
DNA per well (ng)	1327	423		1750
	1327		423	1750
	1327			1750

Note: One day after transfection of the indicated plasmids, IDLV^{d} particles were added at an MOI of 8 VP/cell.

Supplementary Table S5. Experimental scheme corresponding to **Figure 3E** (**Protocol A**)

DONOR: Plasmid^{d}	3.25 ×10^{5} HER.TLR$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)					
	PEI (1mg/ml)5.8 µl per well; Ratio DNA / PEI equivalents = 6					
Reagents	Cas9	gNT (Ctrl)	gTLR.1	gTLR.	Plasmid^{d}	Total (ng)
Construct length (bp)	9551	3056	3046	3046	6194	
DNA per well (ng)	890	284			577	1751
	890		284		577	1751
	890			284	577	1751

Supplementary Table S6. Experimental scheme corresponding to **Figure 3F** (**Protocol B**)

DONOR: Plasmid[d]	3.25 ×10^5 HER.TLR$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)					
	PEI (1mg/ml) 9.6 µl per well; Ratio DNA / PEI equivalents = 10					
Reagents	Cas9	gNT (Ctrl)	gTLR.1	gTLR.2	Plasmid[d]	Total
Construct length (bp)	9551	3056	3046	3046	6194	(ng)
DNA per well (ng)	890	284			577	1751
	890		284		577	1751
	890			284	577	1751

Supplementary Table S7. Experimental scheme corresponding to **Figure 5**

DONOR: pTHG.Donor (Exp.1)	2.0 ×10^5 HEK.EGFP$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)				
	PEI (1mg/ml) 6.2 µl per well; Ratio DNA / PEI equivalents = 9				
Reagents	eCas9	gEGFP	gNT (Ctrl)	pTHG.Donor	Total
Construct length (bp)	9360	3046	3056	3561	(ng)
DNA per well (ng)	733	238		279	1250
	733		238	279	1250
DONOR: pTHG.Donor (Exp.2)	2.0 ×10^5 HEK.EGFP$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)				
	PEI (1mg/ml) 6.2 µl per well; Ratio DNA / PEI equivalents = 9				
Reagents	eCas9.2	gEGFP	gNT (Ctrl)	pTHG.Donor	Total
Construct length (bp)	9403	3046	3056	3561	(ng)
DNA per well (ng)	733	238		279	1250
	733		238	279	1250

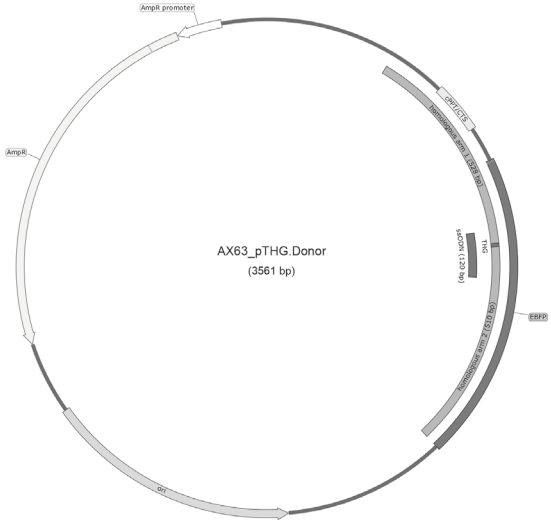
Supplementary Table S8. Experimental scheme corresponding to **Figure 6C**

DONOR: ODN.s / ODN.as	2.5 ×10^5 HEK.EGFP$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)						
	PEI (1mg/ml) 6.2 µl per well; Ratio DNA / PEI equivalents = 9						
Reagents	Cas9	gNT (Ctrl)	gEGFP	ODN.s	ODN.as	Total	Molar ratios
Construct length (bp)	9551	3056	3046	120	120	(ng)	
DNA per well (ng)	642	205		403		1250	1:1:50
	766		244	240		1250	1:1:25
	642		205	403		1250	1:1:50
	642	205			403	1250	1:1:50
	766		244		240	1250	1:1:25
	642		205		403	1250	1:1:50

Supplementary Table S9. Experimental scheme corresponding to **Figure 6E**

DONOR: ODN.s / ODN.as	2.5 ×10^5 HEK.EGFP$^{TetO.KRAB}$ cells per well of 24-well plates (500 µl medium per well with or without Dox)					
	PEI (1mg/ml): 6.2 µl per well; Ratio DNA / PEI equivalents = 9					
Reagents	Cas9	gNT (Ctrl)	gEGFP	ODN.as	Total (ng)	Molar ratios
Construct length (bp)	9551	3056	3046	120		
DNA per well (ng)	642	205		403	1250	1:1:50
	766		244	240	1250	1:1:25
	642		205	403	1250	1:1:50
	553		176	521	1250	1:1:75

Supplementary Notes



>AX63_pTHG.Donor

GGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTACCCCTCACTAAAGGGAACAAAGCTGGTACGAGGACAGGCTGGAGC-
CATGGGCATGGCTACTCAAGCTGATTTGATGGAGTTGGACATGGCCATGGCTGGTGACCACGTCGTGGAATGCCTTCGAATTCAG-
CACCTGCACATGGGACGTCGACCTGAGGTAATTATAACCCGGGCCCTATATATGGATCCAATTGCAATGATCATCATGACAGATCTGCG-
CGCGATCGATATCAGCGCTTTAAATTTGCGCATGCTAGCTATAGTTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTA-
CAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTAAATTTAATTAATCTCGACGGTATCGGTTA-
ACTTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAATTTAAAGAAT-
TACAAAAACAAATTACAAAAATTCAAAATTTTATCGATCACGAGACTAGCCTCGAGGTTTAAACTACGGGATCCAGGCCTAAGCTTACG-
CGTCCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC-
GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTG-
CACCACCGGCAAGCT[GCCCGTGCCCTGGCCCACCCTCGTGACCACCCTG]**ACACATGGC**[GTGCAGTGCTTCAGCCGCTACCCCGAC-
CACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT]CCAGGAGCGCACCATCTTCTTCAAGGACGACGG-
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAG-
GACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG-
CATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC-
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT-
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAGAGCTCGAGAAGTACTAGTG-
GCCACGTGGGCCGTGCACCTTAAGCTTTTAAATAAGGAGGAATAACATATGACCATGATTACGCCAAGCTCCAATTCGCCCTATAGT-
GAGTCGTATTACAATTCACTGGCCGTCGTTTTACTATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG-

CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTA-
ATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG-
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC-
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA-
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC-
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG-
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGT-
GGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG-
TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCT-
CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTAT-
CAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGT-
TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAAC-
TACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATA-
AACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTA-
GAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATG-
GCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC-
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTA-
AGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA-
CACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCT-
TACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGT-
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATAT-
TATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA-
CATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC-
GTCTTCAAGAATT

Map and nucleotide sequence of pTHG.Donor for HDR-mediated editing of *EGFP* into *EBFP*. DNA sequences sharing identity to the target sequence in HEK.EGFP$^{TetO.KRAB}$ cells are indicated in orange; AmpR, β-lactamase ampicillin resistance gene; ori, high-copy number ColE1 prokaryotic origin of replication; cPPT/CTS, central polypurine tract and central termination sequence of HIV-1. As reference, the nucleotide sequences corresponding to the EBFP flurochrome (Thr-His-Gly) and the ssODNs are highlighted in bold and underlined, respectively.

Methods

Cells

The human embryonic retinoblasts HER.TLR$^{TetO.KRAB}$ and their control *TetO*-negative counterparts HER.TLR^{KRAB}, were generated and cultured as detailed elsewhere.[6] Likewise for the human embryonic kidney cells HEK.EGFP$^{TetO.KRAB}$.[6] The HEK293T cells (American Type Culture Collection) used for the generation of IDLV^{d} preparations were maintained in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific). The cells used in this study were mycoplasma free and were kept at 37°C in a humidified-air 10% CO_2 atmosphere.

Recombinant DNA

The gRNA acceptor construct S7_pUC.U6.sgRNA.BveI-stuffer contains a human U6 RNA Pol III promoter and terminator sequence for gRNA expression.[6] The gRNA expression plasmids; Z42_pgTLR.1, Z44_pgTLR.2, AM51_pgNT and AX03_pgEGFP were generated by ligating the annealed oligonucleotide pairs listed in **Supplementary Table S1** into BveI-digested S7_pUC.U6.sgRNA.BveI-stuffer. The plasmid hCas9 was used for expressing the *Streptococcus pyogenes* Cas9 nuclease (Addgene plasmid #41815).[11] The sequence and annotated map of construct AX63_pTHG.donor used for HDR-mediated editing of *EGFP* into *EBFP*, are shown in **Sup-**

plementary Notes. The Addgene plasmid #31475 pCVL SFFV d14 GFP ,[12] herein named Plasmid[d], served as a source of donor DNA in the gene editing experiments performed on HER.TLR$^{TetO.KRAB}$ and HER.TLR^{KRAB} cells. Plasmid[d] is a lentiviral vector construct that harbours the *TLR*-targeting donor template *EGFPtrunc*.[12]

DNA transfections

The DNA transfections performed on cultures of HER.TLR$^{TetO.KRAB}$ were initiated by adding 1 mg/ml of linear 25 kDa polyethyleneimine (PEI, Polysciences) to the different plasmid mixtures diluted in 50 µl of 150 mM NaCl (**Supplementary Tables S2-S6**). These cell cultures were pre-incubated for 10 days in medium lacking or containing doxycycline (Dox) at a final concentration of 0.5 µg/ml. An approximately 10-sec period of vigorous vortexing followed the addition of the PEI polycation to each of the DNA mixtures. Next, the DNA-PEI complexes were let to be formed for 15 min at room temperature after which they were directly added to the culture medium of the various target cells seeded one day before in wells of 24-well plates (Greiner Bio-One). The different transfection mixtures were substituted 6-8 hours later by regular culture medium with or without Dox. At 3 days post-transfection, the cells were sub-cultured every 3-4 days for a period of 10 days and the frequencies of EGFP- and mCherry-positive cells in the cultures containing Dox were determined by flow cytometry (**Supplementary Figure S1**). To activate transgene expression, the cultures initially lacking Dox were exposed to Dox (0.5 µg/ml) for 10 days, after which the frequencies of EGFP- and mCherry-positive cells were also determined in these cultures by flow cytometry. The experimental design, transfection protocols and Dox regimens applied to *TetO*-negative HER.TLR^{KRAB} cells were the same as those applied to HER.TLR$^{TetO.KRAB}$ cells (**Supplementary Figure S1**).

The DNA transfections carried out on cultures of HEK.EGFP$^{TetO.KRAB}$ cells started by adding 1 mg/ml of PEI to the different plasmid mixtures diluted in 50 µl of 150 mM NaCl (**Supplementary Tables S7-S9**). These cell cultures were pre-incubated for 7 days in medium lacking or supplemented with Dox at a final concentration of 0.2 µg/ml. After the addition of PEI to the DNA solutions, an approximately 10-sec period of vigorous vortexing followed. Subsequently, the DNA-PEI complexes were assembled for 15 min at room temperature after which they were directly added to the culture medium of the various target cells that had been seeded one day before in wells of 24-well plates (Greiner Bio-One). The various transfection mixtures were replaced 6-8 hours later by regular culture medium with or without Dox. At 3 days post-transfection, the cells were sub-cultured every 3-4 days for a period of 7 days and the frequencies of EBFP-positive and EGFP-negative cells in the cultures containing Dox were determined by flow cytometry. To activate transgene expression, the cultures that initially had not received Dox were incubated in the presence of Dox (0.2 µg/ml) for an additional 7-day period, after which the frequencies of EBFP-positive and EGFP-negative cells were also determined in these cultures by flow cytometry (**Supplementary Figure S1**).

IDLV production and titration

The assembly of IDLV[d] particles was carried out by transient transfections of HEK293T cells

with lentiviral vector construct Plasmid^{d},[12] together with packaging plasmid AM16_psPAX2.IN^{D116N},[13] and vesicular stomatitis virus glycoprotein-G-pseudotyping construct pLP/VSVG (Thermo Fisher Scientific), as detailed previously.[13, 14] The protocols for the concentration and purification of IDLV^{d} particles released into the producer-cell culture medium were equally detailed elsewhere.[13, 14] Finally, the physical particle titers of the resulting IDLV^{d} stocks were determined by measuring the HIV-1 p24gag antigen with the aid of the RETRO-TEK HIV-1 p24 ELISA kit following the manufacturer's instructions (Gentaur Molecular Products).

Gene editing experiments with single-stranded ODNs

The 120 nucleotide-long single-stranded ODNs ODN.s (5'-GCCCGTGCCCT-GGCCCACCCTCGTGACCACCCTGACACATGGCGTGCAGTGCT-TCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCT-TCAAGTCCGCCATGCCCGAAGGCTACGT-3') and ODN.as (5'-ACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCG-GGGTAGCGGCTGAAGCACTGCACGCCATGTGTCAGGGTGGTCACGAGGGTGGGC-CAGGGCACGGGC-3') were custom synthesized and HPLC-purified (Eurofins Scientific). These ODNs were reconstituted in a solution of 10 mM Tris-Cl and 1 mM EDTA (pH 8.0) to a concentration of 100 pmol/µl. A fifty-fold dilution of this stock was divided in aliquots and stored at -20°C prior to transfection. The ODNs were transfected together with RGN-encoding plasmids into HEK.EGFP$^{TetO.KRAB}$ cells cultured in the absence or in the presence of Dox (0.2 µg/ml) using the previously described PEI-based protocol and the DNA mixtures detailed in **Supplementary Tables S8** and **S9**.

Flow cytometry

The measurements of EGFP-positive, EGFP-negative, EBFP-positive and mCherry-positive cells were performed using a BD LSR II flow cytometer (BD Biosciences). The data were analysed with the support of FlowJo 10.1 software (Tree Star) or BD FACSDiva 6.1.3 software (BD Biosciences). Mock-transfected cells served for establishing background fluorescence thresholds. At least 40,000 viable single cells were analysed per sample.

Statistical analysis

The comparison of the indicated data sets resulting from independent experiments (biological replicates done in different days) were analysed by applying two-tailed Student's *t*-tests ($P<0.05$ considered significant). The GraphPad Prism 6 software package was used for this analysis.

Acknowledgments

The authors thank Rob Hoeben and Ignazio Maggio (Leiden University Medical Center, Departments of Molecular Cell Biology and Pediatrics, respectively) for their critical reading of the manuscript. This work was partially supported by the Dutch Prinses Beatrix Spierfonds (W.OR11–18) and ProQR Therapeutics (Leiden, the Netherlands). X.C. holds a Ph.D. research grant from the China Scholarship Council-Leiden University Joint Scholarship Programme.

Conflict of interest statement. None declared.

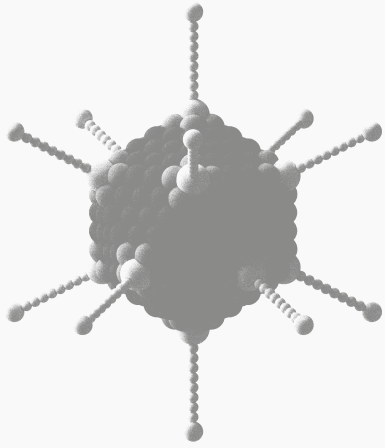
Reference

1. Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014;15:321-334. DOI: 10.1038/nrg3686
2. Maggio I, Goncalves MA. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 2015;33:280-291. DOI: 10.1016/j.tibtech.2015.02.011
3. Chang HHY, Pannunzio NR, Adachi N et al. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017;18:495-506. DOI: 10.1038/nrm.2017.48
4. Heyer WD. Regulation of recombination and genomic maintenance. Cold Spring Harb Perspect Biol. 2015;7:a016501. DOI: 10.1101/cshperspect.a016501
5. Kouzarides T. Chromatin modifications and their function. Cell 2007;128:693-705. DOI: 10.1016/j.cell.2007.02.005
6. Chen X, Rinsma M, Janssen JM et al. Probing the impact of chromatin conformation on genome editing tools. Nucleic Acids Res. 2016;44:6482-6492. DOI: 10.1093/nar/gkw524
7. Daer RM, Cutts JP, Brafman DA et al. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. ACS Synth Biol. 2017;6:428-438. DOI: 10.1021/acssynbio.5b00299
8. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science 2014;346. DOI: 10.1126/science.1258096
9. Wanisch K, Yanez-Munoz RJ. Integration-deficient lentiviral vectors: a slow coming of age. Mol Ther. 2009;17:1316-1332. DOI: 10.1038/mt.2009.122
10. Anders C, Niewoehner O, Duerst A et al. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 2014;513:569-573. DOI: 10.1038/nature13579
11. Mali P, Yang L, Esvelt KM et al. RNA-Guided Human Genome Engineering via Cas9. Science 2013;339:823-826. DOI: 10.1126/science.1232033
12. Certo MT, Ryu BY, Annis JE et al. Tracking genome engineering outcome at individual DNA breakpoints. Nat Methods. 2011;8:671-676. DOI: 10.1038/nmeth.1648
13. Pelascini LPL, Janssen JM, Gonçalves MAFV. Histone Deacetylase Inhibition Activates Transgene Expression from Integration-Defective Lentiviral Vectors in Dividing and Non-Dividing Cells. Hum Gene Ther. 2013;24:78-96. DOI: 10.1089/hum.2012.069
14. Pelascini LP, Goncalves MA. Lentiviral vectors encoding zinc-finger nucleases specific for the model target locus HPRT1. Methods Mol Biol. 2014;1114:181-199. DOI: 10.1007/978-1-62703-761-7_12
15. Urrutia R. KRAB-containing zinc-finger repressor proteins. Genome Biol. 2003;4:231. DOI: 10.1186/gb-2003-4-10-231
16. Iyengar S, Farnham PJ. KAP1 Protein: An Enigmatic Master Regulator of the Genome. J Biol Chem. 2011;286:26267-26276. DOI: 10.1074/jbc.R111.252569
17. Groner AC, Meylan S, Ciuffi A et al. KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. PLoS Genet.

2010;6:e1000869. DOI: 10.1371/journal.pgen.1000869
18. Kuscu C, Arslan S, Singh R et al. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotech. 2014;32:677-683. DOI: 10.1038/nbt.2916
19. Wu X, Scott DA, Kriz AJ et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotech. 2014;32:670-676. DOI: 10.1038/nbt.2889
20. O'Geen H, Henry IM, Bhakta MS et al. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 2015;43:3389-3404. DOI: 10.1093/nar/gkv137
21. Richardson CD, Ray GJ, DeWitt MA et al. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotech. 2016;34:339-344. DOI: 10.1038/nbt.3481
22. Watts FZ. Repair of DNA Double-Strand Breaks in Heterochromatin. Biomolecules 2016;6:47. DOI: 10.1042/BST20110631
23. Garvin AJ, Densham RM, Blair-Reid SA et al. The deSUMOylase SENP7 promotes chromatin relaxation for homologous recombination DNA repair. EMBO Rep. 2013;14:975-983. DOI: 10.1038/embor.2013.141
24. Murr R, Loizou JI, Yang YG et al. Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. Nat Cell Biol. 2006;8:91-99. DOI: 10.1038/ncb1343
25. Baldeyron C, Soria G, Roche D et al. HP1alpha recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair. J Cell Biol. 2011;193:81-95. DOI: 10.1083/jcb.201101030
26. Lee YH, Kuo CY, Stark JM et al. HP1 promotes tumor suppressor BRCA1 functions during the DNA damage response. Nucleic Acids Res. 2013;41:5784-5798. DOI: 10.1093/nar/gkt231
27. Tsouroula K, Furst A, Rogier M et al. Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. Mol Cell. 2016;63:293-305. DOI: 10.1016/j.molcel.2016.06.002
28. Goodarzi AA, Noon AT, Deckbar D et al. ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. Mol Cell. 2008;31:167-177. DOI: 10.1016/j.molcel.2008.05.017
29. Chu VT, Weber T, Wefers B et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotech. 2015;33:543-548. DOI: 10.1038/nbt.3198
30. Robert F, Barbeau M, Ethier S et al. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. 2015;7:93. DOI: 10.1186/s13073-015-0215-6
31. Chylinski K, Makarova KS, Charpentier E et al. Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res. 2014;42:6091-6105. DOI: 10.1093/nar/gku241
32. Burstein D, Harrington LB, Strutt SC et al. New CRISPR–Cas systems from uncultivated microbes. Nature 2017;542:237-241. DOI: 10.1038/nature21059

Chapter 6

Engineered **Viruses** as **Genome Editing** Devices.



Molecular Therapy 24:447-457 (2016).
Chen X, Gonçalves MA.

This image is licensed under the Creative Commons Attribution-Share Alike 4.0 International license.

Abstract

Genome editing based on sequence-specific designer nucleases, also known as programmable nucleases, seeks to modify in a targeted and precise manner the genetic information content of living cells. Delivering into cells designer nucleases alone or together with donor DNA templates, which serve as surrogate homologous recombination (HR) substrates, can result in gene knockouts or gene knock-ins, respectively. As engineered replication-defective viruses, viral vectors are having an increasingly important role as delivery vehicles for donor DNA templates and designer nucleases, namely, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated Cas9 (CRISPR/Cas9) nucleases, also known as RNA-guided nucleases (RGNs). We review this dual role played by engineered viral particles on genome editing while focusing on their main scaffolds, consisting of lentiviruses, adeno-associated viruses, and adenoviruses. In addition, the coverage of the growing body of research on the repurposing of viral vectors as delivery systems for genome editing tools is complemented with information regarding their main characteristics, pros, and cons. Finally, this information is framed by a concise description of the chief principles, tools, and applications of the genome editing field as a whole.

Introduction

Genome editing based on sequence-specific designer nucleases, also known as, programmable nucleases (**Figure 1**) is opening a vast array of scientific and technological possibilities. Its broad range of action stems from granting researchers the means to modify, in a targeted and precise manner, the genetic make-up of cells from an increasing number of higher eukaryotes, including those of humans and other mammals.[1,2,3] In general, this is achieved by inducing double-stranded DNA breaks (DSBs) at predefined chromosomal sequences after designer nuclease delivery into target cells. The delivery of designer nucleases alone (**Figure 2**) or together with so-called donor DNA (**Figure 3**) can result in different targeted genome modification outcomes, each of which resulting from the repair of site-specific DSBs by nonhomologous end-joining (**Figure 2**) or homologous recombination (HR) (**Figure 3**), respectively.

Therefore, a crucial aspect pertaining to the application of genome editing strategies is that of introducing into target cells designer nucleases (**Figure 2**) and, whenever new genetic information needs to be added, surrogate HR substrates in the form of exogenous donor DNA templates (**Figure 3**). Viral vectors are particularly suitable options to introducing genome editing reagents into target cells because, while being replication-defective, they retain the efficient cell entry mechanisms evolved by their wild-type counterparts.[2,4,5] Indeed, as engineered replication-deficient viruses,

viral vectors have been extensively used in academia and industry to deliver foreign genetic payloads into virtually any cell type of interest. Moreover, besides nucleic acids they are also starting to be adapted for the direct transduction of recombinant proteins into target cells, including designer nucleases. In these cases, designer nucleases are fused to structural components of vector particles (for a recent review, see ref. [6]).

The on-going adaptation of viral vectors to genome editing settings builds upon

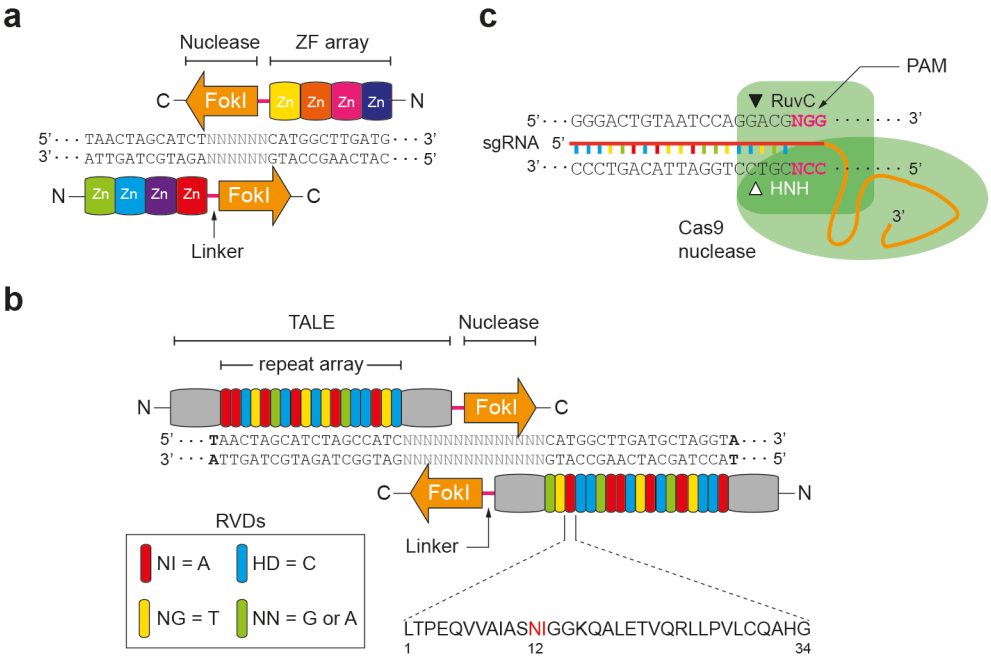


Figure 1

Diagrams of the three principal designer nuclease platforms. (**a**) Zinc-finger nucleases (ZFNs). ZFNs are based on artificial zinc-finger motifs in which two cysteine residues in a β-sheet hairpin and two histidines in an α-helix are tetrahedrally coordinated by a zinc ion. ZF, zinc-finger motif dictating the interaction with a specific nucleotide triplet; FokI, nuclease domain of the type IIS restriction enzyme FokI. (**b**) Transcription activator-like effector nucleases (TALENs). TALENs are based on type III secretory systems of phytopathogenic bacteria (e.g., *Xanthomonas sp.*). TALE, transcription activator-like effector comprising a central DNA-binding domain consisting of an highly conserved 33–34 residue-long repetitive motif; FokI, nuclease domain of the type IIS restriction enzyme FokI; RVDs, repeat variable di-residues located at positions 12 and 13 of each TALE repeat governing the interaction with a particular nucleotide. (**c**) RNA-guided nucleases (RGNs). RGNs are based on prokaryotic type II clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated (Cas) systems (e.g., Cas9 from *Streptococcus pyogenes*). sgRNA, chimeric single-guide RNA consisting of a fusion between a sequence-tailored CRISPR RNA (crRNA) and a scaffolding trans-activating crRNA (tracrRNA). PAM, protospacer adjacent motif (NGG, in the case of *S. pyogenes*); RuvC and HNH, the two nuclease domains of the Cas9 nuclease protein.

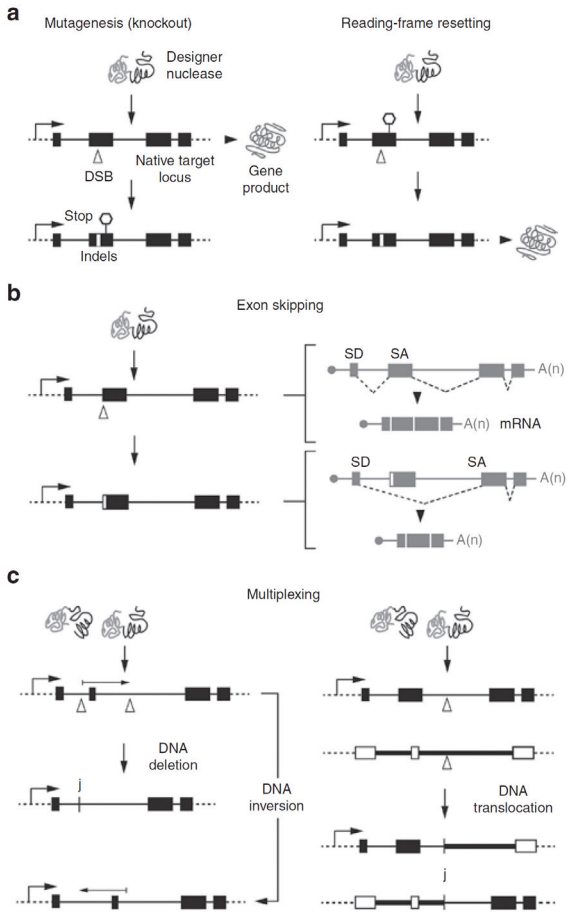


Figure 2

Genome-editing approaches based on nonhomologous end-joining (NHEJ)–mediated repair of chromosomal DSBs using designer nucleases. NHEJ DNA repair mechanisms acting at site-specifically generated DSBs can result in different genome editing outcomes. (**a**) NHEJ-mediated introduction of small insertions and deletions (indels) at the target site often leads to DNA sequence frame shifting, which in turn, can yield gene-specific knockouts whenever protein or RNA coding sequences are targeted (left panel). Conversely, targeted DSB-induced frame shifting can restore the proper reading frame usage (right panel). Of note, targeted mutagenesis resulting from the activity of designer nucleases might also generate protein variants whose mode of action involve dominant negative, or positive, effects (not drawn). (**b**) DNA-level exon skipping can be achieved by targeting genomic sequences corresponding to key splice acceptor elements. (**c**) Coordinated DSB formation by designer nuclease pairs (multiplexing) can yield specific deletions or inversions if the target sites are located in a particular chromosome (left panel) or translocations if they are present in different chromosomes (right panel). *Solid boxes* and *horizontal lines*, exons and introns, respectively; DSB, double-stranded DNA break (*open vertical arrowheads*); *Broken arrows*, *cis*-acting gene regulatory elements including promoters/enhancers. Indels, small insertions and deletions (*open vertical bars*); j, chromosomal DNA junctions formed by non-homologous recombination events triggered by designer nuclease multiplexing. SD and SA, splice donor and splice acceptor, respectively.

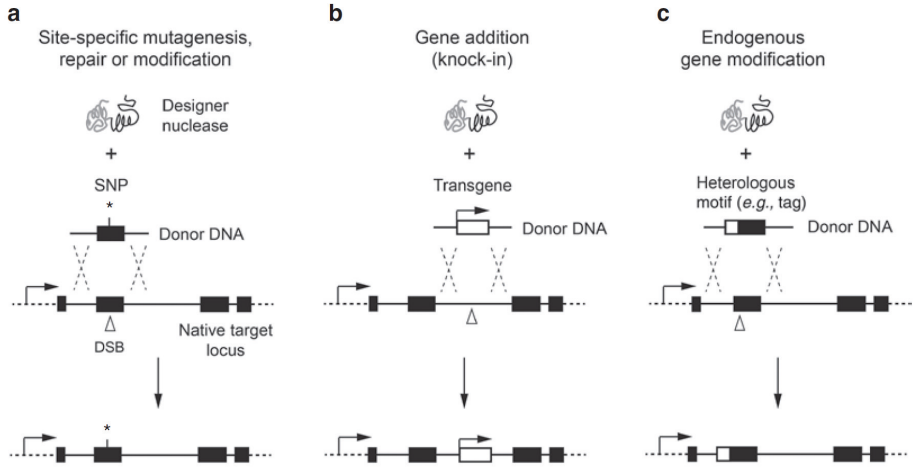


Figure 3

Genome-editing approaches based on HR-mediated repair of chromosomal double-stranded DNA breaks (DSBs) using donor DNA templates and designer nucleases. Robust homology-directed gene targeting, or knock-in, is achievable by combining designer nuclease-induced DSB formation with the delivery of donor DNA templates whose sequences share identity to regions flanking the targeted chromosomal lesion. These genome-editing procedures, based on the activation and recruitment of the homologous recombination (HR) DNA repair machinery, are particularly useful for the precise modification of predefined chromosomal target sequences of choice. Indeed, these DNA-modifying approaches offer the possibility to (**a**) repair, mutate or modify endogenous genes in a site-specific manner, (**b**) introduce entire recombinant transcriptional units (transgenes) into predefined positions in the genome, such as at so-called "safe harbours", and (**c**) endow endogenous proteins with new domains or tag endogenous genes, such as for tracing their expression patterns or isolating their encoded products. *Asterisks*, nucleotide change, such as, addition or deletion of a single nucleotide polymorphism (SNP) or mutation; *Large and small open boxes*, recombinant transcriptional unit (transgene) and heterologous motif (e.g., a tag), respectively. For a description of the other symbols and abbreviations see the legend of **Figure 2**.

a vast amount of knowledge gained from their development for "classical" gene therapy or gene replacement approaches in which the delivered foreign nucleic acids remain mostly in an episomal state or integrate randomly or semi-randomly throughout the target cell's genome.[2,4,5]

Clearly, inserting instead transgenes, or any exogenous DNA for that matter, into specific genomic sequences reduces the chance for various problematic events sometimes emergent whenever using systems that lead to the uncontrolled chromosomal integration of foreign nucleic acids (e.g., retroviral vectors and transposons/transposases). These unwarranted outcomes include positional-effect variegation, transgene silencing and, in some cases, insertional mutagenesis caused by transcriptional deregulation or physical disruption of endogenous target-cell genes.[7] The more defined genome modification outcomes resulting from the aforementioned designer nuclease-assisted genome editing strategies, are having a clear impact in

many fields. For instance, in functional genomics by helping deciphering the role of *cis*- and *trans*-acting nucleotide sequences, in transgenesis by speeding-up animal model generation via direct zygote engineering and in disease modeling by mimicking the origins of certain cancers through the deliberate induction of specific mutations or oncogenic rearrangements. Likewise related to disease modeling, and besides its potential role in future cell therapies, the integration of genome editing with induced pluripotent stem cell (iPSC) technologies is already helping in establishing genotype–phenotype relationships underlying not only monogenic but also polygenic or complex illnesses.[8]

In addition, genome editing strategies are being investigated for developing new treatment modalities aiming at tackling infectious diseases and advancing gene- and cell-based therapies. A first example of the former already exists in the shape of clinical studies testing whether designer nuclease-induced knockout of the HIV-1 coreceptor gene *CCR5* confers therapeutic benefit to acquired immune deficiency syndrome patients.[9,10] In parallel, the investigation of many other candidate gene therapies based on designer nuclease-induced gene knockout and gene knock-in approaches proceeds at the experimental and preclinical levels. These "genome surgery" research lines include deploying designer nucleases for disrupting alleles linked to dominant disorders and triggering homology-directed DNA targeting for repairing or complementing defective genes. The former entails the direct *in situ* correction of endogenous loci; the latter encompasses the targeted insertion of therapeutic DNA at ectopic "safe harbour" loci such as the *AAVS1* (19q13.42). Transgene insertion at such loci results in much higher probabilities for stable and homogeneous expression levels while lessening the chances for the deregulation of target-cell endogenous genes.[11]

In view of the many common goals and substantial overlap between "classical" gene therapy and therapeutic gene-editing research, the co-option of viral vector technologies for the latter purpose is logical and multifaceted in that they are being investigated for delivering not only designer nucleases but also donor DNA templates. Related to this, different types of viral vectors are, in some cases, combined in individual gene-editing transduction protocols. Here we review the roles that the main classes of viral vectors are having on improving the performance of and expanding the scope for genome-editing technologies.

Viral Vectors as Gene-Editing Tools

Lentiviral vectors

Conventional lentiviral vectors based on HIV-1 establish permanent genetic modification of target cells owing to the fact that their integrase-dependent mechanisms ensure semirandom chromosomal insertion of the transported foreign nucleic acids.[12] In "classical" gene therapy settings, these mechanisms are a crucial feature for achieving stable complementation of genetic defects in proliferating target cells and effector progenies.[13] In the context of genome editing approaches, however, the lentiviral DNA insertion mechanisms should best be disabled in order to ensure that the resulting episomal vector templates are available as substrates for HR or for transient designer nuclease expression. As previously mentioned, the short-term presence of designer nucleases in target cells is important for reducing the chances that deleterious effects caused by off-target activity arise. Therefore, by using trans-complementing packaging constructs harboring specific point-mutations in the HIV-1 *pol* region, researchers can assemble lentiviral particles whose integrase moiety contains disabling amino acid substitutions at crucial positions within its catalytic pocket (i.e., D64, D116, and/or E152) (**Figure 4**).[12,14,15] Importantly, these so-called class I integrase mutations are nonpleiotropic in that they interfere specifically with proviral establishment and not with any other of the viral transduction steps, such as, receptor binding, uncoating and nuclear import of the reverse-transcribed linear double-stranded vector genomes. Hence, integrase-defective lentiviral vectors (IDLVs), made with the aid of such packaging constructs serve as valuable vehicles for delivering nucleic acid templates for gene targeting and/or transient designer nuclease gene expression. Of note, similarly to their integration-proficient counterparts, the tropism of IDLV particles are normally altered by endowing them with envelop proteins derived from viruses whose cell surface receptors are different from those engaged by HIV-1. Accordingly, such pseudotyping manoeuvres permit narrowing or expanding the range of cell types transduced by vector particles.[16] For instance, to confer broad host range and high physical particle stability to lentiviral vectors, the vesicular stomatitis virus glycoprotein-G (VSV-G) is often selected as the heterologous envelop moiety (**Figure 4**).

IDLVs were the first viral vectors to be tested in the context of designer nuclease-assisted genome editing experiments.[17] These experiments, involving ZFN technology and various human target cell types (e.g., K562 erythromyeloblastoid leukemia cells, lymphoblastoid cells, and embryonic stem cells), provided an initial proof-of-concept for using IDLVs in designer nuclease-induced gene addition and gene repair studies. The former and latter experiments comprised, respectively, inserting recombinant DNA at specific genomic sequences (i.e., *CCR5*) and correcting *IL2RG* mutations underpinning X-linked severe combined immunodeficiency (X-SCID).

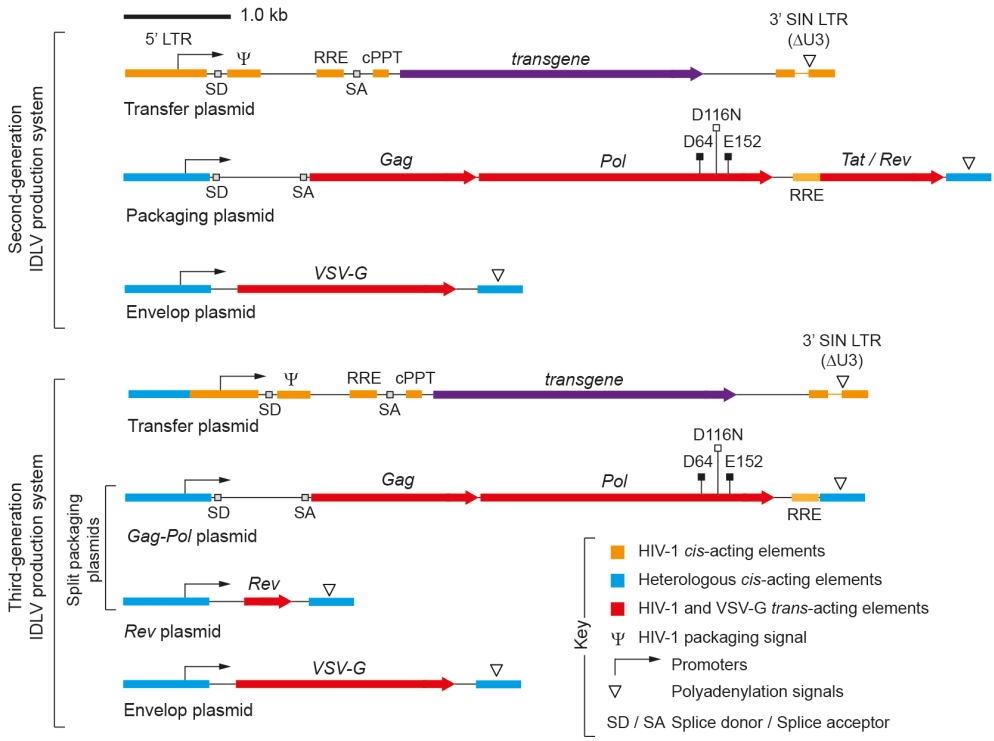


Figure 4

Schematics of the main HIV-1-based vector systems. HIV-1-based vectors are assembled by cotransfecting producer cells (e.g., HEK293T cells) with transfer, envelop, and packaging constructs with the resulting particles being collected and purified after budding from producer cells. Transfer plasmids harbor foreign nucleic acid sequences flanked by HIV-1 5′ and 3′ long terminal repeats (LTRs). To confer HIV-1 Tat independence to vector genome expression, third-generation transfer plasmids have hybrid 5′ LTRs composed of HIV-1 and heterologous transcriptional elements (e.g., *cytomegalovirus* and *Rous sarcoma virus*). Self-inactivating (SIN) vectors are deleted from specific LTR enhancers (ΔU3) to abrogate unwarranted transcriptional activity. Additional HIV-1 *cis*-acting elements include the packaging signal (Ψ), for vector genome encapsidation, the Rev-responsive element (RRE), for nuclear export of unspliced and singly spliced transcripts, and the central polypurine tract (cPPT), for transduction enhancement. The *in trans*-acting envelope plasmids typically encode the pseudotyping VSV-G moiety to confer a pantropic host range to vector particles. The also *in trans*-acting packaging constructs drive expression of HIV-1 *Gag* and *Pol* alone (third-generation) or together with *Tat* and *Rev* (second-generation). Owing to the *Tat* removal and the splitting of Rev from *Gag-Pol* templates, the former systems display a superior biosafety profile. The proteolytic processing of precursor *Pol* sequences yields mature reverse transcriptase and integrase (IN) molecules. Crucially, integrase-defective lentiviral vectors are assembled by using packaging constructs encoding IN moieties with substitutions of one or more amino acids of the DDE triad; D64, D116, and E152. These so-called class I mutations (e.g., D116N) abrogate specifically proviral establishment.

These data revealed that IDLV genomes can serve as efficient HR substrates yielding, in some cell types, homology-directed DNA targeting frequencies exceeding 10% of the total target cell population with the majority of cells harboring mono-allelic insertions. These initial data has been followed-up by various other studies in which IDLV transfer of donor DNA templates resulted in the addition of reporter and therapeutically relevant transgenes into "safe harbour" loci in a diverse set of target cells, including human myocytes,[18,19] human epithelial stem cells,[20] and iPSC lines.[21] Examples of these experiments are the site-specific chromosomal insertion of *microdystrophin* and *FANCA* transgenes into the "safe harbour" *CCR5* locus in, respectively, human muscle progenitor cells[18] and iPSCs from reprogrammed fibroblasts of Fanconi anemia patients.[21] It is worth mentioning however that, in common with any other HR-based genome editing approaches, the recruitment of IDLV donor DNA for gene addition or for gene repair purposes is limited in non-dividing or quiescent cells due to the fact that HR occurs preferentially during the G2/S phase of the cell cycle, when endogenous repairing templates are available.[22] Hence, the cellular DNA of quiescent primary cells, of which many display a high therapeutic relevance, is particularly difficult to edit through HR. An outstanding example of such cells is provided by primitive CD34+ human hematopoietic stem cells (HSCs). HSCs are defined as cells capable of long-term multilineage repopulation of the hematopoietic compartment in conditioned immune-deficient mice.[23] Of note, only genome modification at the HSC level is expected to ensure life-long correction of genetic disorders affecting components of the hematopoietic system. Aiming at improving HR-based genome editing of these cells, Genovese et al.[24] have developed a protocol in which donor DNA and ZFN delivery into HSCs is carried out by IDLV transduction and synthetic mRNA electroporation, respectively. Crucially, this transduction/electroporation protocol is combined with exposing target cells not only to cytokines but also to 16,16-dimethyl-prostaglandin E2 (dmPGE2) mixed with the aryl-hydrocarbon receptor protein antagonist, StemRegenin 1 (SR1). The rationale for including dmPGE2 and SR1 was to interfere with the loss of stem cell properties resulting from HSC exposure to extended *ex vivo* culture conditions and cell cycle-activating cytokines. By using these methods, the authors report that homology-directed gene targeting frequencies at *AAVS1* and *IL2RG* in *bona fide* HSCs are increased, as stringently demonstrated by serial transplantation of human CD34+ cells from primary to secondary NSG (NOD-SCID-*Il2rg*$^{-/-}$) mice. In a subsequent study, Hoban et al.[25] have also tested an *ex vivo* protocol based on the transfer of ZFN-encoding mRNA and IDLV donor templates into bone marrow-derived CD34+ cells for correcting the A-to-T transversion in *β-globin* alleles causing sickle cell anemia.

Hitherto, the delivery of designer nucleases by IDLVs remains mostly restricted to ZFNs. Considering that the yields of functional lentiviral vector particles decrease sharply with increasing transgene size,[26] it is possible that generating IDLVs contain-

ing the 4.1-kb *Cas9* open reading frame (ORF) from *Streptococcus pyogenes* results in IDLV preparations with low functional particle titers. In addition, experimental results indicate that the genetic instability resulting from frequent reverse transcriptase template switching events within TALE repeats in lentiviral vector genomes leads to defective particles.[27] This makes the assembly of TALEN-encoding IDLVs dependent on substantial ORF optimization for minimizing the frequency and length of unstable repetitive tracts.[28] Of note, the same principle of sequence identity reduction has permitted to package and deliver transcriptional units encoding two ZFN monomers in single IDLV particles.[29,30] This two-in-one approach is especially useful at low transduction rates since it ensures that each transduced cell is exposed to both members of a working ZFN pair at the proper 1:1 stoichiometry.

An issue pertaining to the optimal use of IDLVs as designer nuclease expression platforms is that of the susceptibility of their genomes to epigenetic silencing mechanisms in transduced cells.[31,32,33] These mechanisms involve the action of cellular histone deacetylases and have been shown to curtail DSB-induced targeted mutagenesis after IDLV-mediated transfer of ZFN expression units.[34] Finally, another issue regards the susceptibility of free-ended double-stranded IDLV genomes to "illegitimate" recombination processes such as nonhomologous end-joining. As a result, IDLV templates can become "captured" at off-target or spontaneous DSBs and form undesirable DNA structures such as concatemers and non-HR-derived junctions involving target or off-target sequences.[17,19] These adverse genome-modifying events contribute to reduce the fidelity of the genome editing process as a whole.[2]

Adeno-associated viral vectors

In contrast to lentiviral vectors, recombinant adeno-associated viral vectors (rAAVs) lack an integration machinery (**Figure 5**).[35,36] As a result, once in target cell nuclei, their genomes remain mostly in an episomal status with only a small fraction of them becoming incorporated in the cellular DNA (0.1–0.5 integrations per infectious unit)[37] presumably upon nonhomologous end-joining–mediated repair of sporadic chromosomal DSBs.[36] These vectors entered the scene of homology-directed gene targeting during the late 1990's, after the demonstration that viral particle transduction of single-stranded rAAV donor DNA yields more than 1,000-fold higher frequencies of gene repair (up to 1% of the total target-cell population) when compared to those achieved by transfecting conventional donor plasmids.[38] Despite the feasibility of this approach, including in *in vivo* settings,[39] the dominance of off-target insertions combined with the high dependency on large multiplicities of infection (>10^4 total vector particles per cell) and cell selection schemes,[40,41] has contributed to the initiation of research lines based on designer nuclease-assisted rAAV donor DNA targeting.

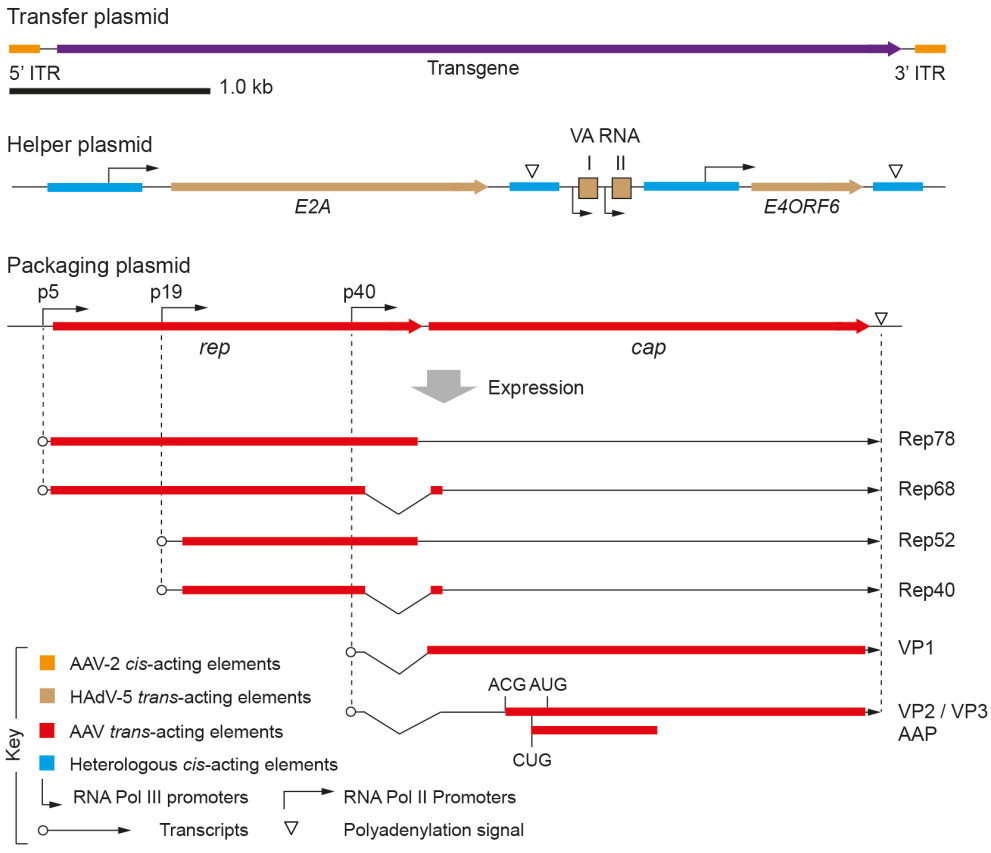


Figure 5

Schematics of the main recombinant adeno-associated viral vector (rAAV) production system. Recombinant AAV vectors are typically assembled by co-transfecting producer cells (e.g., HEK293T cells) with transfer, packaging and helper constructs with the resulting particles being purified after producer cell lysis. Transfer plasmids harbor foreign nucleic acid DNA flanked by 145 bp-long palindromic AAV-2 inverted terminal repeats (ITRs) whose primary sequence and T-shaped secondary structure form the origins of replication. The *in trans*-acting packaging plasmids contain the three AAV ORFs, *rep*, *cap*, and *AAP*. Transcription of *rep* from two different promoters and splicing of each of the resulting transcripts yields four proteins. Rep78 and Rep68 participate in DNA replication; Rep52 and Rep40 are necessary for DNA packaging into preformed empty capsids. The *cap* ORF is transcribed from a single promoter with alternative splicing resulting in two mRNA templates for the synthesis of three viral capsomers (VP1, VP2, and VP3). VP2 and VP3 share the same mRNA template with the former being translated from a "weak" start codon (ACG) located upstream of that for the latter product (AUG). Moreover, another "weak" start codon (CUG) present within the VP2-VP3 mRNA marks the beginning of a third reading frame, which codes for the assembly-activating protein (AAP). Productive wild-type AAV infections depend on the presence of an unrelated virus for providing AAV helper gene functions (*e.g.*, HAdV-5 *E1A-E1B*, *E2A*, *VAI-VAII*, and *E4ORF6*). In the context of rAAV production, these functions are supplied by transfecting E1A-E1B-expressing cells (e.g., HEK293) with a helper construct containing *E2A*, *VAI-VAII* and *E4ORF6*. Often, the packaging and helper functions are combined in a single plasmid.

Like previous data had shown for HR substrates delivered in the context of standard plasmids,[42,43] experiments based on inducing DSBs at chromosomally integrated reporter genes by the homing endonuclease I-SceI, provided a proof-of-concept for combining sequence-specific nucleases with rAAV donor DNA in gene-targeting settings. Indeed, these initial studies revealed that rAAV-based gene targeting can be enhanced by approximately 100-fold if a DSB is generated at a predefined target locus.[44,45] In this realm, and similarly to IDLVs, rAAVs have been mostly used so far for delivering donor DNA templates and ZFNs. Of note, when compared to those of *Cas9* and *TALEN*, *ZFN* ORFs are the smallest (i.e., ~1.2 kb per monomer versus ~4.1 kb and ~3 kb for *S. pyogenes Cas9* and *TALEN* ORFs, respectively). This permits the flexible construction and packaging of transcriptional units encoding one or even two ZFNs in single rAAV particles[46] whose effective maximum capacity is only ~4.5 kb (**Figure 5**). Clearly, in addition to TALEN and Cas9 nuclease delivery, the low packaging capacity of rAAV also introduces some limitations on the designing of HR substrates for the purpose of site-specific addition of whole transcriptional units. In any case, the combination of ZFN and rAAV technologies has clearly proven its potential for not only targeted gene disruption and deletion[46] but also for gene repair strategies. In what the latter genome editing approaches are concerned, these experiments involved the targeting of both reporter and endogenous loci after the delivery of ZFNs and gene correcting templates into a diverse panel of human cell types. These different cell types included, U2OS osteosarcoma cells,[47,48] HEK 293 cells,[46] HeLa cervix carcinoma cells,[48] HT-1080 fibrosarcoma cells,[48] and *bona fide* human embryonic stem cells (ESCs) as well as iPSCs.[49] Noticeably, due to the very diverse range of tools, experimental models and conditions, the gene-targeting frequencies in both absolute and relative terms (i.e., targeted versus random insertion events), varied substantially. As an example, Asuri et al.[49] compared ZFN-induced gene repair levels after transducing ESCs with a HR template packaged either in natural or variant AAV capsids. The latter capsid type, isolated by sequential cycles of biopanning of libraries of cap-mutant viruses on target cells, confers high-level rAAV transduction of hard-to-transfect ESCs and iPSCs. The authors showed, by using a highly quantitative readout system based on the rescue of defective reporter gene expression, that the transfer of corrective donor DNA by the molecularly evolved rAAV variant (R459G) yielded significantly higher (~10-fold) ZFN-induced gene repair levels in ESCs (~1.3% of the total target cell population) when compared to those resulting from using a prototypic, serotype 2-based, rAAV. Importantly, the proportion of random rAAV DNA chromosomal insertions was not augmented by the presence of active ZFNs in the transduced cells. Collectively, this and the above-mentioned studies established that site-specific DSB formation serves as a potent trigger for homology-directed gene targeting of donor DNA delivered in the context of single-stranded rAAV genomes.

Owing to a favorable set of characteristics, rAAVs are particularly suited for testing genome-editing strategies *in vivo*. These characteristics include low immunogenicity in immunocompetent animal models and amenability to tissue tropism modification methodologies based on engineered capsids generated by rational or directed evolution approaches.[50] Moreover, reminiscent of the above-described tropism engineering strategies involving enveloped lentiviral vectors; nonenveloped rAAVs can also be pseudotyped. In this case, rAAV genomes consisting of foreign DNA flanked by prototypic AAV serotype 2 inverted terminal repeats, are packaged within the capsids of other natural AAV isolates such as those of serotypes 1, 5, 6, 8, or 9.[50] These novel capsid-modified rAAVs are powerful gene delivery tools in that they can bypass pre-existing immunity associated with the presence of neutralizing antibodies against particular rAAV serotype(s) and can overcome transductional blocks linked to the absence of viral receptor(s) on the surface of specific cell types or tissues. In addition to the previously mentioned work in which a molecularly evolved rAAV was used,[49] another case in point is provided by the body-wide transduction of murine tissues by rAAV2/6 vectors, that is, AAV serotype 2-derived rAAV genomes pseudotyped by packaging in AAV serotype 6 capsids.[51] Moreover, it has been shown that rAAV2/8 particles achieve frequencies of murine liver cell transduction that are 10- to 100-fold higher than those obtained by using vectors based on other serotypes.[52] Importantly, these experiments equally revealed that the rAAV2/8 gene delivery activity was not hindered in animals preimmunized by exposure to other AAV serotypes.

The relevance and utility of rAAVs in *in vivo* settings is also underscored by the fact that a first proof-of-principle for designer nuclease-induced genome editing *in vivo* involved the use of these vectors in a murine model of hemophilia B, a blood coagulation disorder caused by factor IX deficiency.[53] In particular, rAAV2/8 particles containing a corrective cDNA spanning exons 2 through 8 of human *F9* were administered to new-born hemophilia B mice together with rAAV2/8 particles encoding donor-matched ZFNs targeting intron-1 of a defective human *F9* transgenic allele. Gene targeting was detected and meaningful in that it resulted in 3–7% of normal levels of circulating factor IX that led to the improvement of the disease phenotype as measured by clot-formation kinetic assays. Of note, molecular analysis of genomic DNA from treated mice revealed that therapeutic construct insertions at the intended target site occurred through both homologous and non-HR.[53] The latter, vector genome capture events, were likely caused by end-to-end nonhomologous end-joining of broken chromosome and AAV inverted terminal repeat sequences. A subsequent study extended these findings of AAV/ZFN-mediated *in vivo* therapeutic genome editing to adult hemophilia B mice.[54]

The *in vitro* and *in vivo* transfer of RGN components by rAAVs, has also been initiated. After constructing and validating shortened expression units encoding Cas9 and

sgRNAs, Senís et al.[55] were able to demonstrate delivery of Cas9 alone or together with a sgRNA by single vector particles built on chimeric AAV-DJ capsids. The latter "all-in-one" rAAV construction achieved approximately 8% indel formation at a target miRNA locus in HEK 293T cells when applied at a multiplicity of infection of 10^6 particles per cell. However, in mouse livers, RGN-induced indel formation at the conserved miRNA target locus by different rAAV constructs was invariably below 1% at 2 weeks postadministration. These *in vivo* results have been complemented by other animal model experiments in which rAAV-mediated delivery of RGN components served as a direct, transgenesis-free, approach for studying gene function in the mammalian brain.[56] These initial studies together with the advent of shorter Cas9 variants *bode well* for the implementation of rAAV/RGN tools in different *in vitro* and *in vivo* systems. Indeed, Ran et al.[57] have recently used a comparative genomic analysis to isolate and characterize a *Staphylococcus aureus* Cas9 protein whose relatively small size permits flexible rAAV design, including copackaging of both RGN components within single vector particles. The delivery of these tools into the livers of C57BL/6 mice by rAAV2/8 particles led 1 week after intravenous administration to approximately 5 and 40% indel formation at *Apob* and *Pcsk9* sequences, respectively.[57]

Adenoviral vectors

The sizable packaging capacity of adenoviral vectors (AdVs) combined with their high-titers and efficiency in transducing dividing and nondividing cells, makes them a broadly applicable option for *in vitro* and *in vivo* delivery of designer nucleases and donor DNA templates (**Figure 6**). Similarly to rAAVs, AdVs started to be deployed in the context of homology-directed gene targeting experiments that did not involve designer nuclease-induced DSB formation. In these experiments, helper-dependent AdVs, also known as "gutless" AdVs, were chosen owing to their lack of viral genes, permitting the use of high multiplicities of infection, and high capacity, allowing for large donor DNA packaging and delivery. Indeed, Ohbayashi et al.[58] utilized helper-dependent AdVs with 18.6 kb homology arms to correct a mutation in *HPRT* through HR without the involvement of artificial DSB formation in mouse ES cells. With the emergence of iPSCs, helper-dependent AdVs were also shown to be useful for correcting disease-related mutations in these pluripotent stem cells. In particular, they were used to repair several mutations in *LMNA* alleles associated with laminopathies, thus expanding the application of this gene delivery system to human disease modeling and targeted gene repair.[59] A follow-up study by Aizawa et al.[60], demonstrated that regardless of the transcriptional status of the target gene, helper-dependent AdVs can mediate both gene knock-ins and gene knockouts by HR with high fidelity in both iPSCs and ESCs of human origin. Of note, however, the absolute gene targeting levels achieved by helper-dependent AdVs are rather low requiring as a result the use of drug-based selection pressure for isolating the

desired targeted clones.

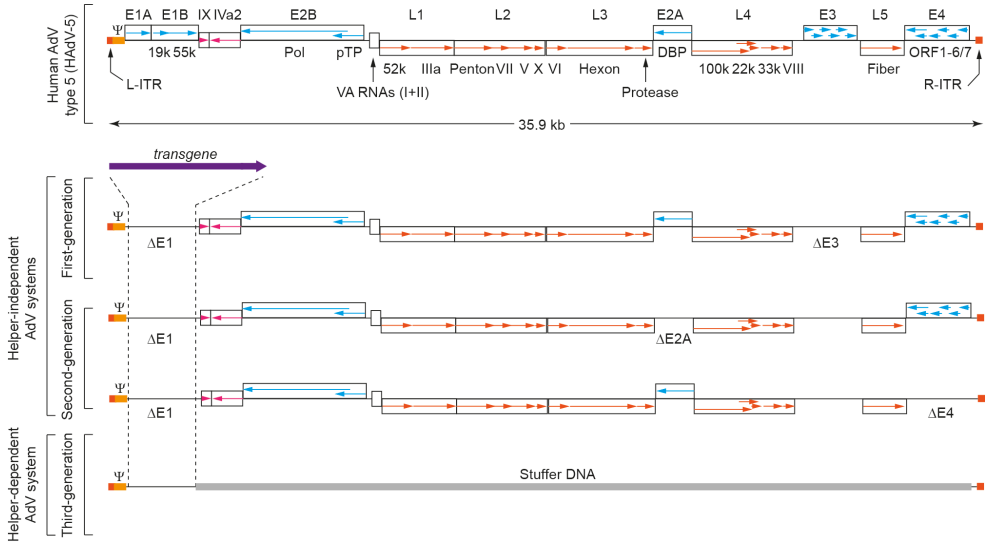


Figure 6

Schematics of the principal adenoviral vector (AdV) systems. The genome structures of the main AdV classes are drawn in relation to that of the prototypic HAdV-5 from species C. The 103 bp-long "left" and "right" inverted terminal repeats (L-ITR and R-ITR, respectively) contain the origins of replication, with the viral DNA packaging signal (Psi) being located adjacent to the L-ITR. The early (E) and late (L) regions are expressed before and after the onset of viral DNA replication, respectively. The former regions (i.e., *E1A-E2A*, *E2A-E2B*, *E3*, and *E4*) encode proteins involved in gene regulation (viral and host) and viral DNA replication; the later encode gene products primarily responsible for virion maturation and assembly (L1-L5). Expression units corresponding to small RNAs (VAI-VAII) and intermediate gene products (IX and IVa2) are also shown. First-generation AdVs lack E1A-E1B or E1A-E1B plus E3. Since E3 is dispensable during *in vitro* replication, all these vectors can be produced in packaging cell lines expressing exclusively the E1 functions (e.g., HEK293 or PER.C6). Second-generation AdVs have deletions in additional early regions (e.g., E2A and/or E4) being, as a result, produced in their respective complementing cell lines. Third-generation AdVs (also known as "gutless" or high-capacity) lack all viral DNA sequences except for the *cis*-acting ITRs and packaging signal. These vectors are produced in E1-complementing cells in the presence of a first-generation helper AdV which furnishes *in trans* all the viral gene products necessary for the replication and assembly of "gutless" AdV particles. The helper has its packaging elements framed by target sites for a site-specific recombinase (e.g., Cre or FLP) so that in recombinase-expressing producer cells is rendered packaging-defective in a selective manner.

Similarly to lentiviral and adeno-associated viral vector systems, AdVs are equally amenable to tropism modification and Good Manufacturing Practice methodologies. The former strategies include exchanging the apical fiber motifs of prototypic species C serotypes, which interact with the Coxsackie B virus and adenovirus receptor (CAR), with those of other natural serotypes (e.g., species B adenoviruses), which interact with other primary receptors. This "fiber swapping" genetic retar-

geting strategy allows by-passing the absence of CAR on the surface of human cells with scientific and therapeutic value such as hematopoietic stem/progenitor cells,[61,62] mesenchymal stromal cells,[63,64,65] and muscle progenitor cells.[65,66] Alternative AdV retargeting methods include capsid modifications by genetic fusion of fiber or pIX capsid proteins to heterologous ligands[67] or by chemical binding of capsid components to targeting moieties.[68] In this regard, it is noteworthy mentioning that the first testing of a therapeutic approach based on genome editing entails ZFN-mediated *CCR5* knockout in CD4+ T-cells from acquired immune deficiency syndrome patients after their *ex vivo* transduction with fiber-modified AdV particles.[9,10] Examples of other genome-editing studies based on the integration of AdV and ZFN technologies include the targeted mutagenesis of endogenous T-cell receptor genes in lymphocytes[69] and of *CCR5* and *β-globin* alleles in hematopoietic stem/progenitor cells.[70,71] Moreover, homology-directed gene targeting induced after AdV-mediated delivery of ZFNs, is equally being pursued in various cell types such as myoblasts, epithelial stem cells, and keratinocytes.[18,20]

Highlighting their versatility, AdV systems have in addition to ZFNs been validated for delivering TALENs and RGN complexes into human somatic cells regardless of their transformation status.[27,72] Concerning the former research it was found that, in striking contrast to lentiviral vector systems, the direct repeat arrays coding for the DNA-binding domains of TALENs are stable during AdV production in complementing packaging cell lines.[27,73] Importantly, the resulting vector preparations led to dose-dependent and high-level (up to 67%) targeted DSB formation in exposed cells (e.g., muscle progenitor cells and mesenchymal stromal cells). The genetic stability of AdVs is also underscored by the fact that transcriptional units encoding ZFN[9,74] or TALEN dimers can be packaged intact in single vector particles.[71,75] Due to the sizable length of *TALEN* ORFs (~3.0 kb per monomer), the latter studies deployed the high-capacity "gutless" AdV platform (**Figure 6**). In addition to the aforementioned muscle progenitor cells and mesenchymal stromal cells, the combination of AdV and TALEN technologies has served for inducing site-specific DSB formation in iPSCs[75] as well as in CD34+ cells isolated from G-CSF-mobilized peripheral blood mononuclear cells.[71]

Recently, various research groups started exploiting the efficient transduction of particular murine tissues by AdVs for studying genetic lesions underlying the emergence of specific cancers and, subsequently, modeling their progression *in vivo*. Such approaches based on the direct induction of targeted genomic changes *in vivo* (e.g., mutations, inversions, and translocations) are more expeditious than those based on transgenic mice and mimic more accurately the stochastic mosaicism characteristic of many tumors. For instance, Zhang et al.[76] succeeded in inducing higher rates of *Apc* mutations in the murine liver after tail vein injection of TALEN-encoding AdVs than those achieved after plasmid hydrodynamic injections (33 versus 7–19%, re-

spectively). Maddalo et al.[77] have in turn deployed RGN-encoding AdVs for inducing an approximately 11 Mb chromosomal inversion involving the *Alk* and *Eml4* loci to model the development of non–small-cell lung cancer *in vivo*.

Besides cancer modeling, other experiments sought to mutagenize *Cebpα*[78] and *Pcsk9* (ref. [79]) in murine livers after the administration of AdVs encoding RGN complexes. The former gene is a transcriptional factor involved in the activation of metabolic target genes; the latter is associated with low-density lipoprotein cholesterol levels, with its loss-of-function correlating with reduced risk of coronary heart disease development. Collectively, these experiments strengthened the view that, together with rAAVs, AdVs serve as a valuable platform for introducing designer nucleases *in vivo*. However, with the expansion and finer follow-up of *in vivo* genome editing procedures, one can expect encountering the immunological hurdles identified previously in countless gene transfer studies in animals. These hurdles include the activation of innate and adaptive immune responses against viral particle components and foreign antigens derived from transgenic and, in the case of helper-independent AdVs, viral ORFs.[80-82] Moreover, the long-term presence of designer nucleases in target tissues adds yet another hurdle that needs to be tackled by, for instance, incorporating regulatory devices for minimizing the risks of chromosomal mutations and/or rearrangements.

In addition to the introduction of designer nucleases into target cells, AdVs are also being exploited as a source of donor DNA templates for homology-directed gene editing after site-specific chromosomal DSB formation by ZFNs,[20,74] TALENs,[19,75] and RGNs.[19,75] In this regard, it has been shown that combining designer nucleases (i.e., TALENs and RGNs) and AdV-mediated donor DNA transfer induces homology-directed gene targeting that is more specific and accurate than that resulting from delivering donor DNA templates through conventional nonviral vectors or IDLVs.[19] The finding of precise genome editing resulting from designer nuclease-induced AdV donor DNA targeting ("Ad.iting", in short) could be attributed to the capping of linear AdV DNA by the 5′ covalently-attached viral terminal protein which, presumably, reduces non-HR events. The resulting targeted, single-copy, donor DNA integrants lead to uniform transgene expression in gene-modified cell populations.[19]

A synopsis of the main characteristics of the viral vector systems being repurposed as gene-editing devices is presented in **Table 1**, whereas their principal pros and cons are summarized in **Table 2**. On the basis of this review and on the information gathered in **Table 1** and **Table 2**, there is no evidence for an "ideal" one-fits-all combination of gene delivery and gene-editing tools. Instead, one can put forward the view that a specific arrangement(s) of these tools is best suited to achieve a particular goal.

In conclusion, viral vectors can serve a dual role in genome engineering efforts by delivering into virtually any human cell type, templates for not only designer nuclease expression but also for targeted chromosomal integration of foreign DNA. These features, combined with their well-established production systems and regulatory history build-up, are expected to foster and expand their application in genome editing settings, including in the realm of translational research.

Table 1. Overview of the main viral vector systems being repurposed as gene editing tools.

Main characteristics	IDLV	rAAV	AdV
Parental virus family, Genus (Prototypic element)	*Retroviridae*, Lentivirus (HIV-1)	*Parvoviridae*, Dependovirus (AAV-2)	*Adenoviridae*, Mastadenovirus (HAdV-5)
Particle structure (Shape)	Enveloped Phospholipid bilayer with trimeric spikes (spherical)	Non-enveloped Protein capsid, fibreless (icosahedral)	Non-enveloped Protein capsid with 12 trimeric fibres (icosahedral)
Vector particle size	~120 nm	~20 nm	~90-100 nm
Vector genome structure	HIV *cis*-acting LTRs and packaging signal flanking foreign DNA	AAV *cis*-acting ITRs flanking foreign DNA	HAdV *cis*-acting ITRs and packaging signal flanking foreign DNA and vector backbone
Typical vector assembly schemes in producer cell lines	Transfection of vector DNA, in *trans* complementing (*Gag*, *Pol*, *Rev*) and pseudotyping (*VSV-G*) constructs	Transfection of vector DNA, in *trans* complementing (*rep*, *cap*) and helper (HAdV genes) constructs	Transfection of vector DNA and propagation of assembled particles in complementing cells
Vector genome polymerases	Particle-associated Reverse transcriptase	Cellular DNA polymerases	Virus-encoded DNA polymerase
Packaged genome structure, polarity (Topology)	2× ssRNA, + strand (linear, free-ended)	1× ssDNA, + or - strand (linear, hairpin-capped)	1× dsDNA, ± strands (linear, protein-capped)
Vector particle assembly processes	Packaging of full-length vector genome transcripts	Packaging of ssDNA from hairpin-primed dsDNA intermediates	Packaging of dsDNA from protein-primed dsDNA intermediates
Vector particle cell entry	Receptor-mediated vector envelop/cell plasmalemma fusion	Receptor-mediated endocytosis via clathrin-coated pits	
Vector genome nuclear entry	Active ds cDNA import via a karyophilic pre-integration complex	Remodelled or intact particle entry through the nuclear pore	Docking of remodelled capsids at the nuclear pore, DNA entry
Prevalent genome status in transduced cell nuclei		Episomal	
Prevalent genome topologies in target cell nuclei	Linear: ds cDNA; circular: ds cDNA 1-LTR and 2-LTR forms	Linear: ssDNA and dsDNA; circular: dsDNA single- and multi-copy forms	Linear: dsDNA
Transduction potency (Target cell replication status)		High (dividing and non-dividing)	
Tropism modifications, including pseudotyping		Straightforward	

AAV-2, human adeno-associated virus type 2; cDNA, complementary DNA; ds, double-stranded; HIV-1, human immunodeficiency virus type 1; HAdV-5, human adenovirus type 5; IDLV, integrase-defective lentiviral vector; ITRs, inverted terminal repeats; LTRs, long terminal repeats; rAAV, recombinant adeno-associated viral vector; ss, single-stranded; *vsv-g*, vesicular stomatitis virus glycoprotein-G gene.

Table 2. Overview of the main pros and cons of IDLV, rAAV and AdV systems.

Parameters	IDLV	rAAV	AdV
Production timelines (research-level batch)	< 2 weeks	< 2 weeks	> 1 month
Vector production up-scaling[a]	+	++	+++
Functional vector particle yields	+	++	+++
Vector particle stability	++	+++	++
Viral genes[b]	Absent	Absent	Present or absent
Effective packaging capacity	< 7 kb	~ 4.5 kb	5-37 kb[c]
Compatibility between packaging capacity and nuclease ORF size[d]	++	+	++ or +++
Genetic stability (mutations, rearrangements and deletions)[e]	+++	+	+
Transgene expression kinetics[f]	++	+	+++
Transgene expression levels	+	++	+++
Susceptibility to transgene silencing	++	+	+
Background chromosomal insertion[g]	++	++	+
In vivo Immunogenicity[h]	+	++	+++
In vivo vector neutralizing antibodies[i]	+	+++	+++

[a]Typically, lentiviral vector production involves transient transfection methods.Virtually all of the AdV and some of the rAAV production platforms encompass instead infection of producer cell cultures with rescued viral vector particles. In these cases, the viral gene-deleted vector genomes are complemented *in trans* by either another viral vector and/or by transgenes stably integrated in the producer cell's DNA.[80,83]

[b]Lentiviral vectors, rAAVs, and helper-dependent, "gutless", AdVs can be made without any of their parental virus coding sequences. The latter system requires, however, a coinfecting helper AdV modified in its packaging signal region to provide *in trans* all the replicative and structural elements needed for vector DNA amplification and packaging, respectively.[80]

[c]The AdV packaging capacity is contingent upon the particular system and varies from a lowest value of ~5 kb for E1-deleted, helper-independent, AdVs to the highest value of ~37 kb for "gutless", helper-dependent, AdVs.[80]

[d]The approximate *ZFN* and *TALEN* ORF lengths are 1.2 and 3.0 kb per monomer, respectively. The combined length of the *S. pyogenes Cas9* ORF plus a typical gRNA expression unit is about 4.4 kb.

[e]The error rates of the HIV-1 reverse transcriptase exceeds those of viral dsDNA polymerases by several orders of magnitude.[84] This contributes to make lentiviral vector genomes mutation-prone.

[f]The transgene expression kinetics of rAAVs is particularly slow due to rate-limiting transduction steps, e.g., vector genome uncoating, nuclear entry and ssDNA to dsDNA conversation for transcriptionally active template generation.[35]

[g]IDLV and rAAV genomes are often "captured" at chromosomal DSBs, presumably due to the action of the nonhomologous end-joining pathway. This contributes to their relatively high levels of basal chromosomal DNA integration.[36,85]

[h]Helper-independent AdVs are particularly immunogenic in part because of adaptive immune responses triggered by "leaky" viral gene expression at high vector doses.[80] rAAVs are immunogenic in part because of capsidspecific T-cell activation at high vector doses.[82]

[i]AdVs and rAAVs based on prototypic serotypes 5 and 2, respectively, are particularly affected by pre-existing neutralizing antibodies due to the fact that a high fraction of the population has been exposed to these, or related, viral serotypes.[81,86]

AdV, adenoviral vector; DSB, double-stranded DNA breaks; IDLV, integrase-defective lentiviral vector ; rAAV, recombinant adeno-associated viral vector.

Acknowledgments

We thank Ignazio Maggio, Josephine M. Janssen, Rob C. Hoeben (Department of Molecular Cell Biology, Leiden University Medical Center, the Netherlands), and Otto-Wilhelm Merten (Généthon, France) for their critical reading of the manuscript. X.C. is the recipient of a PhD research fellowship from the China Scholarship Council-Leiden University Joint Scholarship Programme. M.A.F.V. Gonçalves is coinventor in a patent application on the use of capped donor DNA constructs for targeted DSB-induced genome modifications. The funds for the research carried out in our laboratory include grants from the Dutch Prinses Beatrix Spierfonds (W.OR11-18) and the French AFMTéléthon (16621).

References

1. Gaj, T, Gersbach, CA and Barbas, CF 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31: 397–405.
2. Maggio, I and Gonçalves, MA (2015). Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol 33: 280–291.
3. Kim, H and Kim, JS (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet 15: 321–334.
4. Kay, MA, Glorioso, JC and Naldini, L (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. Nat Med 7: 33–40.
5. Gonçalves, MA (2005). A concise peer into the background, initial thoughts and practices of human gene therapy. Bioessays 27: 506–517.
6. Skipper, KA and Mikkelsen, JG (2015). Delivering the goods for genome engineering and editing. Hum Gene Ther 26: 486–497.
7. Biasco, L, Baricordi, C and Aiuti, A (2012). Retroviral integrations in gene therapy trials. Mol Ther 20: 709–716.
8. Sterneckert, JL, Reinhardt, P and Schöler, HR (2014). Investigating human disease using stem cell models. Nat Rev Genet 15: 625–639.
9. Perez, EE, Wang, J, Miller, JC, Jouvenot, Y, Kim, KA, Liu, O et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases.
Nat Biotechnol 26: 808–816.
10. Tebas, D, Stein, WW, Tang, I, Frank, SQ, Wang, G, Lee, SK et al. (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N Engl J Med 370: 901–910.
11. Sadelain, M, Papapetrou, EP and Bushman, FD (2012). Safe harbours for the integration of new DNA in the human genome. Nat Rev Cancer 12: 51–58.
12. Mátrai, J, Chuah, MK and VandenDriessche, T (2010). Recent advances in lentiviral vector development and applications. Mol Ther 18: 477–490.
13. Naldini, L (2011). Ex vivo gene transfer and correction for cell-based therapies. Nat Rev Genet 12: 301–315.
14. Philpott, NJ and Thrasher, AJ (2007). Use of nonintegrating lentiviral vectors for gene therapy. Hum Gene Ther 18: 483–489.
15. Wanisch, K and Yáñez-Muñoz, RJ (2009). Integration-deficient lentiviral vectors: a slow coming of age. Mol Ther 17: 1316–1332.
16. Cronin, J, Zhang, XY and Reiser, J (2005). Altering the tropism of lentiviral vectors through pseudotyping. Curr Gene Ther 5: 387–398.
17. Lombardo, A, Genovese, P, Beausejour, CM, Colleoni, S, Lee, YL, Kim, KA et al. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25: 1298–1306.

18. Benabdallah, BF, Duval, A, Rousseau, J, Chapdelaine, P, Holmes, MC, Haddad, E et al. (2013). Targeted Gene Addition of Microdystrophin in Mice Skeletal Muscle via Human Myoblast Transplantation. Mol Ther Nucleic Acids 2: e68.
19. Holkers, M, Maggio, I, Henriques, SF, Janssen, JM, Cathomen, T and Gonçalves, MA (2014). Adenoviral vector DNA for accurate genome editing with engineered nucleases. Nat Methods 11: 1051–1057.
20. Coluccio, A, Miselli, F, Lombardo, A, Marconi, A, Malagoli Tagliazucchi, G, Gonçalves, MA et al. (2013). Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. Mol Ther 21: 1695–1704.
21. Rio, P, Baños, R, Lombardo, A, Quintana-Bustamante, O, Alvarez, L, Garate, Z et al. (2014). Targeted gene therapy and cell reprogramming in Fanconi anemia. EMBO Mol Med 6: 835–848.
22. Kass, EM and Jasin, M (2010). Collaboration and competition between DNA doublestrand break repair pathways. FEBS Lett 584: 3703–3708.
23. Doulatov, S, Notta, F, Laurenti, E and Dick, JE (2012). Hematopoiesis: a human perspective. Cell Stem Cell 10: 120–136.
24. Genovese, P, Schiroli, G, Escobar, G, Di Tomaso, T, Firrito, C, Calabria, A et al. (2014). Targeted genome editing in human repopulating haematopoietic stem cells. Nature 510: 235–240.
25. Hoban, MD, Cost, GJ, Mendel, MC, Romero, Z, Kaufman, ML, Joglekar, AV et al. (2015). Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. Blood 125: 2597–2604.
26. Kumar, M, Keller, B, Makalou, N and Sutton, RE (2001). Systematic determination of the packaging limit of lentiviral vectors. Hum Gene Ther 12: 1893–1905.
27. Holkers, M, Maggio, I, Liu, J, Janssen, JM, Miselli, F, Mussolino, C et al. (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Res 41: e63.
28. Yang, L, Guell, M, Byrne, S, Yang, JL, De Los Angeles, A, Mali, P et al. (2013). Optimization of scarless human stem cell genome editing. Nucleic Acids Res 41: 9049–9061.
29. Joglekar, AV, Hollis, RP, Kuftinec, G, Senadheera, S, Chan, R and Kohn, DB (2013). Integrase-defective lentiviral vectors as a delivery platform for targeted modification of adenosine deaminase locus. Mol Ther 21: 1705–1717.
30. Abarrategui-Pontes, C, Créneguy, A, Thinard, R, Fine, EJ, Thepenier, V, Fournier, le RL et al. (2014). Codon swapping of zinc finger nucleases confers expression in primary cells and in vivo from a single lentiviral vector. Curr Gene Ther 14: 365–376.
31. Kantor, B, Ma, H, Webster-Cyriaque, J, Monahan, PE and Kafri, T (2009). Epigenetic activation of unintegrated HIV-1 genomes by gut-associated short chain fatty acids and its implications for HIV infection. Proc Natl Acad Sci USA 106: 18786–18791.
32. Pelascini, LP, Janssen, JM and Gonçalves, MA (2013). Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. Hum Gene Ther 24: 78–96.
33. Joglekar, AV, Stein, L, Ho, M, Hoban, MD, Hollis, RP and Kohn, DB (2014). Dissecting the mechanism of histone deacetylase inhibitors to enhance the activity of zinc finger nucleases delivered by integrase-defective lentiviral vectors. Hum Gene Ther 25: 599–608.
34. Pelascini, LP, Maggio, I, Liu, J, Holkers, M, Cathomen, T and Gonçalves, MA (2013). Histone deacetylase inhibition rescues gene knockout levels achieved with integrasedefective lentiviral vectors encoding zinc-finger nucleases. Hum Gene Ther Methods 24: 399–411.
35. Gonçalves, MA (2005). Adeno-associated virus: from defective virus to effective vector. Virol J 2: 43.
36. Miller, DG, Petek, LM and Russell, DW (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. Nat Genet 36: 767–773.

37. McCarty, DM, Young, SM Jr and Samulski, RJ (2004). Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu Rev Genet 38: 819–845.
38. Russell, DW and Hirata, RK (1998). Human gene targeting by viral vectors. Nat Genet 18: 325–330.
39. Miller, DG, Wang, PR, Petek, LM, Hirata, RK, Sands, MS and Russell, DW (2006). Gene targeting in vivo by adeno-associated virus vectors. Nat Biotechnol 24: 1022–1026.
40. Hirata, R, Chamberlain, J, Dong, R and Russell, DW (2002). Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. Nat Biotechnol 20: 735–738.
41. Vasileva, A and Jessberger, R (2005). Precise hit: adeno-associated virus in gene targeting. Nat Rev Microbiol 3: 837–847.
42. Lukacsovich, T, Yang, D and Waldman, AS (1994). Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. Nucleic Acids Res 22: 5649–5657.
43. Rouet, P, Smih, F and Jasin, M (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci USA 91: 6064–6068.
44. Miller, DG, Petek, LM and Russell, DW (2003). Human gene targeting by adenoassociated virus vectors is enhanced by DNA double-strand breaks. Mol Cell Biol 23: 3550–3557.
45. Porteus, MH, Cathomen, T, Weitzman, MD and Baltimore, D (2003). Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. Mol Cell Biol 23: 3558–3565.
46. Ellis, BL, Hirsch, ML, Porter, SN, Samulski, RJ and Porteus, MH (2013). 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.
47. Händel, EM, Gellhaus, K, Khan, K, Bednarski, C, Cornu, TI, Müller-Lerch, F et al. (2012). Versatile and efficient genome editing in human cells by combining zincfinger nucleases with adeno-associated viral vectors. Hum Gene Ther 23: 321–329.
48. Rahman, SH, Bobis-Wozowicz, S, Chatterjee, D, Gellhaus, K, Pars, K, Heilbronn, R et al. (2013). The nontoxic cell cycle modulator indirubin augments transduction of adenoassociated viral vectors and zinc-finger nuclease-mediated gene targeting. Hum Gene Ther 24: 67–77.
49. Asuri, P, Bartel, MA, Vazin, T, Jang, JH, Wong, TB and Schaffer, DV (2012). Directed evolution of adeno-associated virus for enhanced gene delivery and gene targeting in human pluripotent stem cells. Mol Ther 20: 329–338.
50. Kotterman, MA and Schaffer, DV (2014). Engineering adeno-associated viruses for clinical gene therapy. Nat Rev Genet 15: 445–451.
51. Gregorevic, P, Blankinship, MJ, Allen, JM, Crawford, RW, Meuse, L, Miller, DG et al. (2004). Systemic delivery of genes to striated muscles using adeno-associated viral vectors. Nat Med 10: 828–834.
52. Gao, GP, Alvira, MR, Wang, L, Calcedo, R, Johnston, J and Wilson, JM (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. Proc Natl Acad Sci USA 99: 11854–11859.
53. Li, H, Haurigot, V, Doyon, Y, Li, T, Wong, SY, Bhagwat, AS et al. (2011). In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature 475: 217–221.
54. Anguela, XM, Sharma, R, Doyon, Y, Miller, JC, Li, H, Haurigot, V et al. (2013). Robust ZFN-mediated genome editing in adult hemophilic mice. Blood 122: 3283–3287.
55. Senís, E, Fatouros, C, Große, S, Wiedtke, E, Niopek, D, Mueller, AK et al. (2014). CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. Biotechnol J 9: 1402–1412.
56. Swiech, L, Heidenreich, M, Banerjee, A, Habib, N, Li, Y, Trombetta, J et al. (2015). In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. Nat Biotech-

nol 33: 102–106.
57. Ran, FA, Cong, L, Yan, WX, Scott, DA, Gootenberg, JS, Kriz, AJ et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature 520: 186–191.
58. Ohbayashi, F, Balamotis, MA, Kishimoto, A, Aizawa, E, Diaz, A, Hasty, P et al. (2005). Correction of chromosomal mutation and random integration in embryonic stem cells with helper-dependent adenoviral vectors. Proc Natl Acad Sci USA 102: 13628–13633.
59. Liu, GH, Suzuki, K, Qu, J, Sancho-Martinez, I, Yi, F, Li, M et al. (2011). Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. Cell Stem Cell 8: 688–694.
60. Aizawa, E, Hirabayashi, Y, Iwanaga, Y, Suzuki, K, Sakurai, K, Shimoji, M et al. (2012). Efficient and accurate homologous recombination in hESCs and hiPSCs using helperdependent adenoviral vectors. Mol Ther 20: 424–431.
61. Shayakhmetov, DM, 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.
62. Knaän-Shanzer, S, Van Der Velde, I, Havenga, MJ, Lemckert, AA, De Vries, AA 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.
63. Knaän-Shanzer, S, van de Watering, MJ, van der Velde, I, Gonçalves, MA, Valerio, D and de Vries, AA (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.
64. Gonçalves, MA, de Vries, AA, Holkers, M, van de Watering, MJ, van der Velde, I, van Nierop, GP et al. (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.
65. Janssen, JM, Liu, J, Skokan, J, Gonçalves, MA and de Vries, AA (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.
66. Gonçalves, MA, Holkers, M, Cudré-Mauroux, C, van Nierop, GP, Knaän-Shanzer, S, van der Velde, I et al. (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.
67. Vellinga, J, Van der Heijdt, S and Hoeben, RC (2005). The adenovirus capsid: major progress in minor proteins. J Gen Virol 86(Pt 6): 1581–1588.
68. Kreppel, F and Kochanek, S (2008). Modification of adenovirus gene transfer vectors with synthetic polymers: a scientific review and technical guide. Mol Ther 16: 16–29.
69. Provasi, E, Genovese, P, Lombardo, A, Magnani, Z, Liu, PQ, Reik, A et al. (2012). Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. Nat Med 18: 807–815.
70. Li, L, Krymskaya, L, Wang, J, Henley, J, Rao, A, Cao, LF et al. (2013). Genomic editing of the HIV-1 coreceptor CCR5 in adult hematopoietic stem and progenitor cells using zinc finger nucleases. Mol Ther 21: 1259–1269.
71. Saydaminova, K, Ye, X, Wang, H, Richter, M, Ho, M, Chen, H et al. (2015). Efficient genome editing in hematopoietic stem cells with helper-dependent Ad5/35 vectors expressing site-specific endonucleases under microRNA regulation. Mol Ther Methods Clin Dev 1: 14057.
72. Maggio, I, Holkers, M, Liu, J, Janssen, JM, Chen, X and Gonçalves, MA (2014). Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Sci Rep 4: 5105.

73. Holkers, M, Cathomen, T and Gonçalves, MA (2014). Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells. Methods 69: 179–187.
74. Zhang, W, Chen, H, Zheng, X, Wang, D, Ji, H, Xia, H et al. (2014). Targeted genome correction by a single adenoviral vector simultaneously carrying an inducible zinc finger nuclease and a donor template. J Biotechnol 188C: 1–6.
75. Suzuki, K, Yu, C, Qu, J, Li, M, Yao, X, Yuan, T et al. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. Cell Stem Cell 15: 31–36.
76. Zhang, S, Li, L, Kendrick, SL, Gerard, RD and Zhu, H (2014). TALEN-mediated somatic mutagenesis in murine models of cancer. Cancer Res 74: 5311–5321.
77. Maddalo, D, Manchado, E, Concepcion, CP, Bonetti, C, Vidigal, JA, Han, YC et al. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature 516: 423–427.
78. Cheng, R, Peng, J, Yan, Y, Cao, P, Wang, J, Qiu, C et al. (2014). Efficient gene editing in adult mouse livers via adenoviral delivery of CRISPR/Cas9. FEBS Lett 588: 3954–3958.
79. Ding, Q, Strong, A, Patel, KM, Ng, SL, Gosis, BS, Regan, SN et al. (2014). Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. Circ Res 115: 488–492.
80. Gonçalves, MA and de Vries, AA (2006). Adenovirus: from foe to friend. Rev Med Virol 16: 167–186.
81. Louis Jeune, V, Joergensen, JA, Hajjar, RJ and Weber, T (2013). Pre-existing antiadeno-associated virus antibodies as a challenge in AAV gene therapy. Hum Gene Ther Methods 24: 59–67.
82. Mingozzi, F and High, KA (2013). Immune responses to AAV vectors: overcoming barriers to successful gene therapy. Blood 122: 23–36.
83. Kotin, RM (2011). Large-scale recombinant adeno-associated virus production. Hum Mol Genet 20(R1): R2–R6.
84. Duffy, S, Shackelton, LA and Holmes, EC (2008). Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet 9: 267–276.
85. Gabriel, R, Lombardo, A, Arens, A, Miller, JC, Genovese, P, Kaeppel, C et al. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol 29: 816–823.
86. Seregin, SS and Amalfitano, A (2009). Overcoming pre-existing adenovirus immunity by genetic engineering of adenovirus-based vectors. Expert Opin Biol Ther 9: 1521–1531.

Chapter 7

The **Emerging** Role of **Viral** Vectors as Vehicles for ***DMD*** Gene Editing.

Genome Medicine 8:59 (2016).
Maggio I*, Chen X*, Gonçalves MA. (*co-first author)

Abstract

Duchenne muscular dystrophy (DMD) is a genetic disorder caused by mutations in the dystrophin-encoding *DMD* gene. The *DMD* gene, spanning over 2.4 megabases along the short arm of the X chromosome (Xp21.2), is the largest genetic locus known in the human genome. The size of *DMD*, combined with the complexity of the DMD phenotype and the extent of the affected tissues, begs for the development of novel, ideally complementary, therapeutic approaches. Genome editing based on the delivery of sequence-specific programmable nucleases into dystrophin-defective cells has recently enriched the portfolio of potential therapies under investigation. Experiments involving different programmable nuclease platforms and target cell types have established that the application of genome-editing principles to the targeted manipulation of defective *DMD* loci can result in the rescue of dystrophin protein synthesis in gene-edited cells. Looking towards translation into the clinic, these proof-of-principle experiments have been swiftly followed by the conversion of well-established viral vector systems into delivery agents for *DMD* editing. These gene-editing tools consist of zinc-finger nucleases (ZFNs), engineered homing endoculeases (HEs), transcription activator-like effector nucleases (TALENs), and RNA-guided nucleases (RGNs) based on clustered, regularly interspaced, short palindromic repeats (CRISPR)–Cas9 systems. Here, we succinctly review these fast-paced developments and technologies, highlighting their relative merits and potential bottlenecks, when used as part of *in vivo* and *ex vivo* gene-editing strategies.

Background

Duchenne muscular dystrophy (DMD) is a lethal X-linked genetic disorder (affecting approximately 1 in 5000 boys) [1] caused by mutations in the ~2.4-megabase *DMD* gene [2] which lead to irrevocable muscle wasting owing to the absence of dystrophin in the striated muscle cell lineage [3]. Although dystrophin-disrupting mutations can be of different types, 68 % of them consist of intragenic large deletions [4]. These deletions can be found along the entire length of the enormous *DMD* locus, with 66 % nested within a major, recombination-prone, hotspot region spanning exons 45 through 55 [4]. The resulting joining of exons flanking DMD-causing mutations by pre-mRNA splicing yields transcripts harboring out-of-frame sequences and premature stop codons, which are presumably degraded by nonsense-mediated mRNA decay mechanisms.

In muscle cells, the long rod-shaped dystrophin protein anchors the intracellular cytoskeleton to the extracellular matrix via a large glycoprotein complex embedded in the plasma membrane called the dystrophin-associated glycoprotein complex (DGC). This structural link is fundamental for proper cellular signaling and structural integrity. Indeed, in the absence of dystrophin, a relentless degenerative

process is initiated that consists of the substitution of muscle mass by dysfunctional fibrotic and fat tissues [3]. As time elapses, patients with DMD become dependent on a wheelchair for ambulation and, later on, require breathing assistance. Crucially, with the aid of palliative treatments, which include supportive respiratory and cardiac care, the life expectancy of patients with DMD is improving and a greater proportion of these patients now reach their late 30s [5].

Targeting the root cause of DMD

The complexity of DMD, combined with the extent of affected tissue, demands the development of different, ideally complementary, therapeutic approaches. The goal of pursuing parallel approaches is to target different aspects and stages of the disease and hence maximize the length and quality of patients' lives. Towards this end, various candidate therapies are currently under intense investigation [3, 5, 6]. These research lines include: (1) mutation-specific exon skipping via modulation of pre-mRNA splicing by antisense oligonucleotides; (2) compensatory upregulation of dystrophin's autosomal paralog utrophin by small-molecule drugs or artificial transcription factors; (3) cell therapies involving allogenic myogenic stem/progenitor cell transplantation; and (4) gene therapies based on the delivery of shortened versions of dystrophin (for example, microdystrophins) to affected tissues. Of note, these recombinant microdystrophins are devoid of centrally located motifs that are, to some extent, dispensable. The miniaturization bypasses the fact that the full-length 11-kilobase (kb) dystrophin coding sequence is well over the packaging limit of most viral vector systems.

More recently, genome-editing strategies based on sequence-specific programmable nucleases have been proposed as another group of therapies for DMD [7–10]. Programmable nucleases are tailored to induce double-stranded DNA breaks (DSBs) at predefined positions within complex genomes [11–13]. In chronological order of appearance, these enzymes are: zinc-finger nucleases (ZFNs) [14], engineered homing endonucleases (HEs) [15], transcription activator-like effector nucleases (TALENs) [16–18], and RNA-guided nucleases (RGNs) based on dual RNA-programmable clustered, regularly interspaced, short palindromic repeat (CRISPR)–Cas9 systems [19–22] (**Fig. 1**). HEs, also known as meganucleases, from the LAGLIDADG family can be engineered to cleave DNA sequences other than those of their natural target sites. The designing of new substrate specificities depends, however, on complex protein engineering efforts involving the screening of large combinatorial assemblies of HE parts [15]. Regardless, redesigned HE were shown to create indel footprints at intronic *DMD* sequences, albeit at very low frequencies (<1 % of target alleles in human myoblasts) [23]. In contrast to the construction of redesigned HEs, the modular nature of the DNA-binding motifs of ZFNs and TALENs makes them more amenable to protein engineering [14, 16–18]. Of note, the assembly of highly specific TALENs is

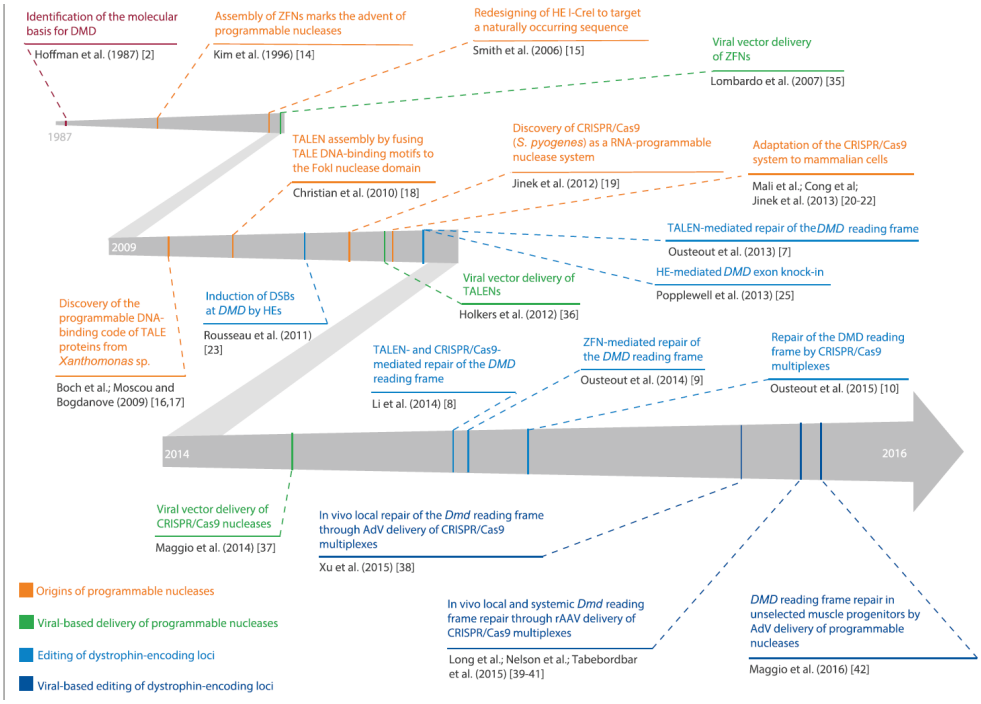


Fig. 1

Milestones on the path towards somatic genetic therapies for Duchenne muscular dystrophy that rely on viral-based *DMD* editing. The time marks correspond to the first release date of the referenced articles (for example, advanced online publication). AdV adenoviral vector, CRISPR–Cas9 clustered regularly interspaced short palindromic repeats-associated Cas9 nuclease, DMD Duchenne muscular dystrophy, DSB double-stranded DNA break, HE homing endonuclease, rAAV recombinant adeno-associated virus, TALE transcription activator-like effector

particularly straightforward owing to a simple one-to-one relationship between the binding of each of their DNA-binding modules, that is, transcription activator-like effectors (TALE) repeats, and a specific nucleotide [16, 17]. Among other features, ZFNs and TALENs differ from RGNs in that they are chimeric enzymes that assemble at their target nucleotide sequences as catalytically active dimers through protein–DNA binding, whereas RGNs are ribonucleoprotein complexes whose DNA cutting specificities are ultimately governed by DNA–RNA hybridization. Indeed, RGNs consist of a Cas9 endonuclease and a sequence-customizable single-guide RNA (sgRNA) moiety that addresses the protein component to a specific target site. Typically, the target site consists of 18–20 nucleotides complementary to the 5′ end of the sgRNA and a protospacer adjacent motif (PAM; NGG and NNGRRT in the case of the prototypic *Streptococcus pyogenes* Cas9 and its smaller orthologue *Staphylococcus aureus* Cas9, respectively) [19, 24]. Hence, in comparison with the strictly protein-based systems, RGNs are more versatile owing to their mode of construction, which does

not involve protein engineering [11–13].

Regardless of the DNA cutting system that is selected, the repair of the ensuing DSBs by different endogenous cellular DNA repair processes can yield specific genome editing outcomes. For example, the engagement of homologous recombination (HR) and non-homologous end joining (NHEJ) mechanisms can result in targeted exogenous DNA additions and endogenous DNA deletions, respectively [11–13]. The incorporation of small insertions and deletions (indels) following the repair of DSBs by NHEJ can also be exploited for knocking out *trans*-acting and *cis*-acting genomic elements [11–13]. By operating at the DNA level, such interventions can potentially lead to the correction of disease-causing mutations on a permanent basis.

DMD gene editing

DMD editing based on targeted addition of "exon patches" corresponding to missing or disrupted coding sequences might become ideal therapeutic options as they result in the synthesis of full-length dystrophin [8, 25]. Proof-of-principle experiments demonstrated that combining *DMD*-repairing exon patches with engineered meganucleases [25], RGNs, or TALENs [8] can indeed restore full-length message coding for dystrophin. At present, however, most *DMD* editing approaches under investigation are based on inducing NHEJ to disrupt or delete specific sequences [7–10]. These strategies exploit the fact that, in contrast to HR, NHEJ is active in both dividing and post-mitotic cells [26, 27], which makes these approaches more amenable to both *ex vivo* and *in vivo* applications (**Table 1**). The NHEJ-based strategies also capitalize on the fact that internally truncated in-frame *DMD* transcripts, despite being shorter than the full-length *DMD* transcript, often yield functional dystrophins [28–30]. Indeed, such dystrophins are characteristic of patients with Becker muscular dystrophy, whose disease phenotypes are milder than those of their counterparts with DMD [28–30]. Therefore, programmable nucleases have been tailored for correcting defective *DMD* alleles by targeting: (1) splicing sites for inducing DNA-borne exon skipping; (2) exonic sequences for resetting reading frames and "overwriting" downstream premature stop codons; and (3) flanking intronic sequences for directly excising mutations through the use of pairs of programmable nucleases (multiplexing) [7–10]. DNA-borne exon skipping by NHEJ-mediated splicing motif knockout and reading-frame resetting by frame shifting are mutation-specific and rely on the fraction of indel footprints that yield in-frame sequences. Importantly, the resulting indels might introduce immunogenic epitopes into *de novo*-synthesized dystrophin molecules. Depending on certain variables (for example, revertant mutation backgrounds), these epitopes might be recognized as foreign by the immune system. Related to this potential issue, T-cell immunity to dystrophin epitopes was detected in two patients undergoing a clinical trial based on recombinant adeno-associated viral vector (rAAV) delivery of a microdystrophin construct [31].

In contrast to those triggering single-exon deletions, the *DMD* correction approaches based on targeted multi-exon deletions do not give rise to indel-derived epitopes and are applicable to a wider range of DMD-causing genotypes, with *de novo*-generated intronic junctions leading to predictable in-frame mRNA templates [10, 32]. However, multiplexing approaches carry increased risks for unwarranted, possibly deleterious, genome-modifying events (for example, off-target DSBs, inversions, and translocations), owing to their dependency on two programmable nucleases rather than one [12]. These increased risks will be present despite the fact that targeted DSBs in boys with DMD will be restricted to a single allele.

Viral-based *DMD* editing

The clinical application of *DMD*-editing concepts will require improved methods for delivering large and complex molecular tools into target cells, as well as increasing the efficiency, specificity, and fidelity of the ensuing DNA modifications [12]. Similarly to their effective contribution to "classic" gene replacement therapies [33], viral vectors are expected to become instrumental tools for investigating and developing therapeutic gene-editing approaches *ex vivo* and *in vivo* (for a recent review on the adaptation and testing of viral vector systems for genome editing purposes, see [34]). Indeed, ZFNs, TALENs, and RGNs have all been shown to be amenable to viral vector delivery [35–37] (**Fig. 1**). More recently, adenoviral vectors (AdVs) and rAAVs have been successfully converted into *DMD*-editing agents in both patient-derived cells and mouse models of DMD [38–42] (**Fig. 1**).

In vivo

The *Dmd*^{mdx} mouse model has a (mild) dystrophic phenotype that is due to a nonsense mutation located in exon 23 of the *Dmd* gene; historically, this has been the principal animal model for investigating DMD-targeted therapies and certain pathophysiological aspects of the disease [43]. In one study, conventional, commonly used, serotype-5 AdVs constructed to encode either *S. pyogenes* Cas9 or sgRNAs that targeted sequences flanking *Dmd* exons 21 through 23 were co-injected into the gastrocnemius muscles of newborn *Dmd*^{mdx} mice [38]. At 3 weeks post-injection, dystrophin synthesis was readily detected in transduced muscle fibers. A semi-quantitative assay based on western blot analysis estimated that these fibers contained ~50 % of the wild-type levels of dystrophin. The gene-edited muscle regions displayed reduced Evans blue dye uptake under rest and force-generating conditions, indicating improved muscle fiber integrity.

A notorious characteristic of prototypic serotype-5 AdVs is their immunogenicity and, although they can be made without viral genes [34, 44], capsid-cell interactions can still trigger strong innate immune responses [45, 46]. In addition, the high prevalence of neutralizing antibodies directed against the capsids of serotype-5 AdVs in the

human population has contributed to spurring the development of AdVs based on alternative serotypes [45]. Historically, these immunological determinants have in fact precluded the efficacious deployment of AdV technologies in "classic" gene therapy settings in which long-term maintenance of transduced cells is a prerequisite. AdVs are currently mostly used in human individuals either as oncolytic or vaccination agents [47]. The use of AdVs in translational *in vivo* gene editing will require dampening their immunogenicity and improving their targeting to specific cell types or organs. These efforts will be heavily guided by insights into the biology of host–vector interactions [45, 46]. For example, while serotype-5 AdVs bind through their fibers to the coxsackievirus and adenovirus receptor (CAR) to enter cells *in vitro* [48], their uptake by liver cells after intravenous administration *in vivo* is CAR-independent and governed by the interaction of their hexons with blood coagulation factors [49].

Three other studies investigated the *in vivo* delivery of RGN components (that is, sgRNA and Cas9 nucleases) by capsid-pseudotyped rAAVs for inducing the in-frame deletion of *Dmd* exon 23. These rAAV particles consist of rAAV DNA from serotype 2 packaged in capsids from AAV serotype 8 (rAAV-8) [40] or serotype 9 (rAAV-9) [39, 41], whose tropism for striated mouse muscle had previously been established [50, 51]. Pairs of these vectors encoding sgRNAs and either *S. pyogenes* Cas9 [39] or the smaller *S. aureus* Cas9 [40, 41] were co-administered into newborn and adult *Dmd*^{mdx} mice. Nelson and colleagues detected abundant dystrophin protein synthesis 8 weeks after co-injecting a mixture of rAAV-8 particles encoding *S. aureus* Cas9 and cognate sgRNAs into tibialis anterior muscles [40]. Importantly, treated muscles had improved contractibility and force-generating functions. Finally, by capitalizing on the well-established performance of rAAV-8 after systemic administration in mice [50], Nelson and colleagues were able to detect dystrophin in cardiac muscle tissue after a single intravenous injection [40].

Instead of rAAV-8, Long and colleagues used rAAV-9 to introduce *S. pyogenes* RGN complexes into striated muscle tissues of newborn *Dmd*^{mdx} mice [39]. Dystrophin was detected in striated muscle tissues after local and systemic administration of the engineered viral vectors [39]. Consistent with the slow kinetics of gene expression from rAAVs, which might in part be related to the processes underlying the conversion of vector DNA from a single-stranded to a transcriptionally active double-stranded form [52], a time-dependent increase in dystrophin buildup was observed. For instance, tibialis anterior muscles of postnatal day 12 *Dmd*^{mdx} mice subjected to direct intramuscular injections with the engineered viral vector contained approximately 8 and 26 % of dystrophin-positive fibers at 3 and 6 weeks post-administration, respectively [39].

In the third study, Tabebordbar and coworkers used rAAV-9 pairs for delivering *S. aureus* Cas9 and sgRNAs to the tibialis anterior muscle of dystrophin-defective *Dmd*-

mdx mice [41]. Similarly to the results of the two other studies on rAAV-mediated *Dmd* exon 23 deletion experiments [39, 40], administration of the rAAV-9 pairs led to robust rescue of dystrophin protein synthesis in transduced muscles and to a concomitant measurable improvement in functional parameters (that is, specific force and force drop) compared with those in unedited controls [41]. In addition, intraperitoneal co-injection of rAAV-9 particles into dystrophic mice led to frequencies of *Dmd* exon 23 excision in cardiac and skeletal muscle tissues ranging from 3 to 18 %, as determined by RT-PCR, depending on the muscle groups analyzed [41]. Importantly, *Dmd*-editing rAAV-9 particles were also administered intramuscularly or systemically to *Pax7-ZsGreen Dmd*^{mdx} mice whose satellite cells are marked by green fluorescence. Subsequently, after isolating, expanding, and inducing myogenic differentiation of the Pax7-ZsGreen-positive cells, the authors reported in-frame *Dmd* exon 23 deletions in myotubes derived from these cells [41]. The population of Pax7-positive satellite cells harbors the resident mononuclear stem cell population of skeletal muscle and is typically lodged between the sarcolemma of muscle fibers and the basal lamina [53]. The "stemness" qualities of self-renewal and lifelong differentiation capacity make these tissue-specific stem cells ideal substrates for regenerative medicine approaches for treating muscular dystrophies as, in contrast to their committed progenitor offspring, these cells support robust long-term tissue homeostasis and repair [54, 55]. Recent experiments in transgenic *Dmd*^{mdx} mice showed that, in addition to its other functions, dystrophin has a transient but critical regulatory role in activated Pax7-positive satellite cells, which further supports the therapeutic relevance of this cell population. In particular, the 427-kilodalton dystrophin isoform is expressed at very high levels in these cells, where it governs asymmetric cell division, a process that is indispensable for maintaining the stem cell pool and for generating committed Myf5-positive myoblast progenitors for muscle repair [56]. Among other processes, this mechanism presumably involves interactions between the spectrin-like repeats R8 and R9 of dystrophin and Mark2, a protein that regulates cell polarity [56, 57]. If conserved in humans, this cell-autonomous mechanism would be evidence that DMD is also a stem cell disease, which would strengthen the view that satellite cells should be preferential targets for DMD therapies, in addition to differentiated cells. Interestingly, the very high amounts of dystrophin seen in activated Pax7-positive satellite cells are followed by very low and intermediate levels of the protein in myoblasts and differentiated muscle cells, respectively [56]. Such differentiation-stage-specific oscillations in dystrophin amounts strengthen the rationale for repairing the genetic defects by direct endogenous *DMD* editing, as this strategy is expected to restore proper regulation of dystrophin synthesis.

Taken together, these findings demonstrate that rAAV delivery of RGN complexes can result in the structural improvement of treated striated tissues and also lead to the partial rescue of specific muscle functions in dystrophic mice. Although dystro-

phin synthesis was detected at 6 months after a single injection in one experiment [40], no long-term detailed assessments of these approaches were done. Regardless, the available data do support the potential of these vectors as *in vivo DMD*-repairing agents, thus warranting further research. Future developments should include assuring the transient presence of programmable nucleases in post-mitotic tissues, preclinical testing in large outbred animal models [43], and identifying or engineering rAAV capsids that have preferential tropism for human striated muscle cells, including satellite cells, while bypassing the host's humoral immunity against prevalent AAV serotypes [58].

The administration of rAAVs to some human individuals resulted in clinical endpoints that had not been predicted on the basis of the available preclinical data. These findings are simultaneously sobering and illuminating. An example is provided by the elimination of transduced hepatocytes in patients with hemophilia B, which was due to the development of a dose-dependent T-cell response to capsid epitopes from an rAAV-2-encoding human factor IX [59]. This type of dose-dependent cellular immune response has also been documented in human skeletal muscle cells transduced with rAAVs [60], although it is of note that the emergence of T-cell responses directed against rAAV capsid epitopes does not always equate with the elimination of transduced muscle cells [61]. In addition, short-term immune suppression might help to dampen cellular immune responses in muscular dystrophy patients subjected to high-doses of rAAV particles [62]. It is worth mentioning, however, that the altered immune cell composition and inflammatory environment that characterize dystrophic muscle tissue might introduce potential confounding factors associated with *in vivo* rAAV delivery. Knowledge about these issues and preclinical data obtained from canine models of DMD [63–65] are guiding the design of new clinical trials based on the administration of rAAVs to patients with DMD [66]. Further insights are being gathered from the application of rAAVs to patients suffering from other muscular disorders such as Limb-girdle muscular dystrophy caused by α-sarcoglycan deficiency [67]. In particular, there is mounting evidence for the importance of restricting transgene expression to muscle cells by using tissue-specific promoters [67]. In the future, muscle-restricted transgene expression might be further improved by combining transcriptional with transductional targeting through rAAVs with capsids with a strict tropism for human muscle tissue. The recently discovered pan-AAV receptor AAVR [68] is likely to have an important role in this research; for instance, by shedding light on rAAV transduction profiles in different cell types, including immune-related cells. Therefore, although rAAVs have a substantially milder immunogenic profile than that of AdVs, they also need to be adapted for translational *in vivo* gene-editing purposes, which, as for AdVs, will be rooted in an increasing knowledge about vector-host interactions and biodistribution at the organismal level. Finally, in the context of future clinical protocols for *in vivo DMD* editing, the synthesis of pro-

grammable nucleases should be restricted not only spatially but also temporally.

Ex vivo

Ex vivo DMD editing strategies to generate genetically corrected human cells with myoregenerative capacity for autologous transplantation can also be envisaged (**Table 1**). These approaches offer a controlled genome-modification environment, bypass vector-neutralizing antibodies, and minimize direct contact between the patient and immunogenic components, such as those from vector particles, gene-editing tools, and allogenic donor cells (**Table 1**). Importantly, provided that clinically applicable delivery vehicles of gene-editing tools become available, *ex vivo DMD* editing can naturally build upon the numerous investigations that are being conducted on the isolation, characterization, and testing of human myogenic cells isolated from different tissues for treating muscular dystrophies [69–73]. These cellular substrates include satellite cells [53, 54] and their committed myoblast progeny [74], induced pluripotent stem cells [75], mesenchymal stromal cells [76, 77], vasculature-associated mesoangioblasts/pericytes [78], and blood-derived CD133+ cells [79]. Of note, the latter two cell types have been shown to be amenable to systemic administration in animal models and, to some extent, can colonize their satellite cell niche [80–82]. In addition, mesoangioblasts/pericytes and CD133+ cells have entered early stage clinical testing in the context of allogenic cell therapies for DMD [83, 84]. These clinical investigations complement earlier and ongoing testing of allogenic myoblast transplantation that are based on intramuscular injections [71–73, 85, 86].

Despite these encouraging developments, the hurdles towards the clinical application of *ex vivo* DMD cell therapies remain numerous and complex. Preeminent examples of such hurdles include achieving sufficient numbers of undifferentiated cells *in vitro*, as well as robust cell engraftment, migration, and differentiation of the transplanted graphs *in vivo*. Ideally, the transplanted cells should also be capable of homing to damaged tissue after systemic administration and should dedifferentiate or transdifferentiate (when belonging to muscle and non-muscle lineages, respectively) into satellite cells (**Table 1**). Therefore, although certain therapeutic-cell candidates are well positioned to fulfil some of these criteria, none of them fulfils all of the criteria yet [69, 72]. For example, CD133+ blood-derived cells and mesoangioblasts/pericytes have been shown to be compatible with systemic administration procedures in preclinical models of muscular dystrophies [78, 79] , but their contribution to effective myoregeneration requires further investigation. In contrast, the features of human satellite cells make them natural, highly potent, muscle-repairing entities. Besides being available in diverse human muscle groups, satellite cells have the capacity to readily engraft as functional stem cells and robustly contribute to *de novo* muscle repair in xenotransplantation experiments [72]. However, harvested satellite cells are not amenable to systemic administration or current *ex vivo* culture

conditions, as they readily differentiate into myoblasts with reduced regenerative capacity [87]. Importantly, the latter hurdle might not be insurmountable, as ongoing research indicates that extrinsic factors such as the composition and elasticity of culture vessels can be modulated to mimic the rigidity of the native satellite cell niche (that is, 12 instead of ~106 kilopascals) and, in doing so, enable the in vitro survival and self-renewal of bona fide satellite cells [88]. The development of such biomimetic tissue-engineering technologies directed to the *in vitro* expansion of human satellite cells is in demand.

In addition to that of skeletal muscle, cardiac muscle impairment is a key component of DMD that also needs to be tackled in future therapies. Despite intense research on the isolation and characterization of stem and progenitor cells for the repair of damaged heart tissue (for example, after ischemia), so far there is no evidence for a significant functional improvement of the myocardium through the cell-autonomous differentiation of the transplanted cells into mature, electrically coupled cardiomyocytes [89, 90].

Other equally important areas for further research in the field of *DMD*-targeted regenerative medicine are: (1) deepening our knowledge about the origins and biology of the various cell therapy candidates and their interaction(s) with their respective niches; (2) gathering all possible information on the fate and behavior of transplanted cells from ongoing and future cell therapy trials; (3) moving forward with gene replacement approaches involving stable transduction of recombinant constructs; and (4) testing different gene-editing reagents and strategies for developing autologous cell transplantation approaches. Regarding the latter research avenue, it will be crucial to efficiently introduce different gene-editing tools into human muscle progenitor cells and non-muscle cells with myogenic capacity. AdVs outperform rAAVs in *ex vivo* settings owing to their higher functional vector particle titers, larger packaging capacity (up to 37 kb), and faster kinetics of transgene expression [34, 52]. Our laboratory has recently reported that tropism-modified AdVs are particularly efficient and versatile vehicles for introducing RGNs and TALENs into CAR-negative myoblasts from patients with DMD [42]. The strict episomal nature of the transduced AdV genomes enabled transient high-level expression of programmable nucleases that corrected native *DMD* alleles and yielded permanent and regulated dystrophin synthesis. In this work, we exploited targeted NHEJ-mediated correction of DMD-causing intragenic deletions by reading-frame resetting, DNA-borne exon skipping, and in-frame excision of single or multiple exons [42]. The rescue of dystrophin synthesis could be readily detected in unselected populations of target cells [42]. Bypassing the need for cell selection expedients is expected to simplify and help translate *ex vivo* *DMD* editing protocols to the clinic. Moreover, AdV-based delivery systems will aid with assessing and comparing different *DMD* editing reagents and strategies in panels of human myogenic cells harboring the various *DMD* mutations, which are not

represented in the currently available animal models. In addition, the well-defined *in vitro* conditions permit the straightforward monitoring of intended as well as unwarranted or potentially deleterious interactions between the gene-editing reagents and the human genome (**Table 1**). Prominent examples of such quality controls will include the genome-wide tracking of adverse DNA-modifying events directly in patient cells (for example, off-target activities of programmable nucleases).

Conclusions and future directions

The application of genome-editing principles for *DMD* repair purposes is expanding the range of genetic therapy options for tackling DMD. In this context, the coopting of viral vector systems as carriers of programmable nucleases is set to have an important role in the path to DNA-targeted DMD therapies and, along the way, in defining the best strategies and optimizing the corresponding reagents. In view of the complexity of the DMD phenotype and the extent of the affected tissues, it is sensible to consider that future DMD therapies will profit from integrating complementary approaches. For example, the simultaneous treatment of skeletal and cardiac tissues from patients with DMD might be approached by combining *ex vivo* and *in vivo* gene-editing strategies, respectively. Such schemes can potentially address the skeletal and heart components of DMD while circumventing the current lack of cell entities capable of differentiating into functional cardiomyocytes. Regardless of the particular therapy or combination of therapies ultimately selected, there is widespread agreement that they should preferably be applied as early as possible so that most striated musculature is still in place and the degeneration process can be halted or, ideally, reversed in the treated muscle groups. Finally, the insights gained from these *DMD*-targeted research efforts will probably also be useful for devising advanced genetic therapies for addressing other neuromuscular disorders for which, at present, there are no therapeutic options available.

✓ Pros × Cons	Viral-based *DMD* editing	
	Ex vivo	*In vivo*
Background	× Knowledge about the grafting of different types of myogenic cells into recipient human muscles is scarce.	✓ Builds upon an increasing amount of knowledge on the *in vivo* administration of viral vectors into recipient human muscles (e.g. microdystrophin-encoding rAAVs).
Production	✓ Potentially less dependent on large-scale viral vector batches. × Reliant on the upscaling of cell culture systems. × The required numbers of certain myogenic cell types might not be achievable owing to senescence (e.g. myoblasts). × The current protocols do not permit culturing *bona fide* skeletal muscle stem cells (i.e. satellite cells) *in vitro.*	✓ Independent from the upscaling of cell culture systems. × More reliant on large-scale viral vector batches.
Delivery	✓ Well-defined genetic modification environment that enables careful monitoring of procedures, events and outcomes. ✓ Lower stringency for monitoring the biodistribution (e.g. gonads and shedding of vector elements) × Protocols for effective myogenic cell engraftment, migration and differentiation need to be improved (e.g. via signaling gradients and cell-autonomous reprogramming of iPSCs). × Local and locoregional administration of myogenic cells might be difficult to apply to a broad range of muscle groups. × Protocols for the systemic delivery and tissue homing of myogenic cells need to be developed.	✓ Direct exposure to gene-editing tools facilitates *in situ* correction of differentiated striated muscle tissues. ✓ Possible *in situ* transduction of resident tissue-specific stem cells might generate a long-term source of gene-edited muscle progenitor cells. ✓ Expanding range of viral vector pseudotypes enables the investigation of different transduction patterns, e.g. tropism for affected tissues while avoiding APCs. Such transductional targeting can easily be complemented with transcriptional targeting (i.e. use of tissue-specific promoters). × Local and locoregional administration of viral vector particles might be difficult to apply to a broad range of muscle groups. × Protocols for the systemic delivery of viral vectors to affected tissues need to be improved. × Higher stringency for monitoring the biodistribution (e.g. gonads and shedding of vector elements.
Strategy	✓ Relies mostly on targeting replicating cells that are amenable to gene-editing approaches based on NHEJ as well as HR.	× Relies mostly on targeting post-mitotic cells, which are less amenable to HR-based gene editing principles.
Immunology	✓ Minimizes the exposure of the patient to immunogenic components of viral vectors and gene-editing tools. ✓ Possibly compatible with the re-administration of gene-edited autologous cells. ✓ Avoids the blocking of viral vector particles by neutralizing antibodies present in the majority of the human population.	× Patient exposure to immunogenic components of vector particles and gene-editing tools. Possible mounting of cellular responses to transduced cells displaying foreign epitopes. × Anti-vector neutralizing antibodies in the majority of the human population. Serotype cross-neutralizing activity might render vector pseudotyping and vector re-administration ineffective.

Table 1. *In vivo* approaches entail the direct administration of gene-editing viral vectors to the patient. *Ex vivo* approaches encompass the *in vitro* transduction of patient-derived cells (for example, myogenic stem or progenitor cells) with gene-editing viral vectors, which is followed by cell culture and autologous transplantation back into the patient. Both treatment modalities can, in principle, be applied either locally or systemically. APCs antigen-presenting cells, HR homologous recombination, iPSCs induced pluripotent stem cells, NHEJ non-homologous end joining, rAAVs recombinant adeno-associated viruses

Abbreviations

AdV	adenoviral vector
APC	antigen-presenting cell
CAR	coxsackievirus and adenovirus receptor
CRISPR	clustered, regularly interspaced, short palindromic repeats
DGC	dystrophin-associated glycoprotein complex
DMD	Duchenne muscular dystrophy
DSB	double-stranded DNA break
HR	homologous recombination
indel	insertion and deletion
iPSC	induced pluripotent stem cell
kb	kilobase
NHEJ	non-homologous end joining
PAM	protospacer adjacent motif
rAAV	recombinant adeno-associated viral vector
RGN	RNA-guided nuclease
sgRNA	single-guide RNA
TALE	transcription activator-like effector
TALEN	transcription activator effector-like nuclease
ZFN	zinc-finger nuclease

Acknowledgments

The authors are grateful to Josephine M. Janssen and Jin Liu for their excellent support and input to the research activities carried out in our research group. The authors also thank Rob Hoeben for his critical reading of the manuscript (Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, the Netherlands). XC is the recipient of a PhD research fellowship from the China Scholarship Council-Leiden University Joint Scholarship Programme. This work was in part funded by grants from the Dutch Prinses Beatrix Spierfonds (W.OR11-18) and the French AFMTéléthon (16621).

References

1. Mendell JR, Shilling C, Leslie ND, Flanigan KM, Al-Dahhak R, Gastier-Foster J, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann Neurol. 2012;71:304–13. doi: 10.1002/ana.23528.
2. Hoffman EP, Brown RH, Jr, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell. 1987;51:919–28. doi: 10.1016/0092-8674(87)90579-4.
3. Guiraud S, Aartsma-Rus A, Vieira NM, Davies KE, van Ommen GJ, Kunkel LM. The pathogenesis and therapy of muscular dystrophies. Annu Rev Genomics Hum Genet. 2015;16:281–308. doi: 10.1146/annurev-genom-090314-025003.
4. Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, et al. The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. Hum Mutat. 2015;36:395–402. doi: 10.1002/humu.22758.
5. Guiraud S, Chen H, Burns DT, Davies KE. Advances in genetic therapeutic strategies for Duchenne muscular dystrophy. Exp Physiol. 2015;100:1458–67. doi: 10.1113/EP085308.

6. Konieczny P, Swiderski K, Chamberlain JS. Gene and cell-mediated therapies for muscular dystrophy. Muscle Nerve. 2013;47:649–63. doi: 10.1002/mus.23738.
7. Ousterout DG, Perez-Pinera P, Thakore PI, Kabadi AM, Brown MT, Qin X, et al. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. Mol Ther. 2013;21:1718–26. doi: 10.1038/mt.2013.111.
8. Li HL, Fujimoto N, Sasakawa N, Shirai S, Ohkame T, Sakuma T, et al. Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. Stem Cell Rep. 2015;4:143–54. doi: 10.1016/j.stemcr.2014.10.013.
9. Ousterout DG, Kabadi AM, Thakore PI, Perez-Pinera P, Brown MT, Majoros WH, et al. Correction of dystrophin expression in cells from Duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. Mol Ther. 2015;23:523–32. doi: 10.1038/mt.2014.234.
10. Ousterout DG, Kabadi AM, Thakore PI, Majoros WH, Reddy TE. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nat Commun. 2015;6:6244. doi: 10.1038/ncomms7244.
11. Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014;15:321–34. doi: 10.1038/nrg3686. [PubMed] [Cross Ref]
12. Maggio I, Gonçalves MA. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 2015;33:280–91. doi: 10.1016/j.tibtech.2015.02.011.
13. Chandrasegaran S, Carroll D. Origins of programmable nucleases for genome engineering. J Mol Biol. 2016;428:963–89. doi: 10.1016/j.jmb.2015.10.014.
14. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 1996;93:1156–60. doi: 10.1073/pnas.93.3.1156.
15. Smith J, Grizot S, Arnould S, Duclert A, Epinat J-C, Chames P, et al. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res. 2006;34:e149. doi: 10.1093/nar/gkl720.
16. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009;326:1509–12. doi: 10.1126/science.1178811.
17. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science. 2009;326:1501. doi: 10.1126/science.1178817.
18. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010;186:757–61. doi: 10.1534/genetics.110.120717.
19. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21. doi: 10.1126/science.1225829.
20. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–6. doi: 10.1126/science.1232033.
21. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–23. doi: 10.1126/science.1231143.
22. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. eLife. 2013;2:e00471. doi: 10.7554/eLife.00471.
23. Rousseau J, Chapdelaine P, Boisvert S, Almeida LP, Corbeil J, Montpetit A, Tremblay JP. Endonucleases: tools to correct the dystrophin gene. J Gene Med. 2011;13:522–37. doi: 10.1002/jgm.1611.
24. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520:186–91. doi: 10.1038/nature14299.

25. Popplewell L, Koo T, Leclerc X, Duclert A, Mamchaoui K, Gouble A, et al. Gene correction of a duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in. Hum Gene Ther. 2013;24:692–701. doi: 10.1089/hum.2013.081.
26. Kass EM, Jasin M. Collaboration and competition between DNA double-strand break repair pathways. FEBS Lett. 2010;584:3703–8. doi: 10.1016/j.febslet.2010.07.057.
27. Rothkamm K, Kruger I, Thompson LH, Lobrich M. Pathways of DNA double-strand break repair during the mammalian cell cycle. Mol Cell Biol. 2003;23:5706–15. doi: 10.1128/MCB.23.16.5706-5715.2003.
28. England SB, Nicholson LV, Johnson MA, Forrest SM, Love DR, Zubrzycka-Gaarn EE, et al. Very mild muscular dystrophy associated with the deletion of 46 % of dystrophin. Nature. 1990;343:180–2. doi: 10.1038/343180a0.
29. Nakamura A, Yoshida K, Fukushima K, Ueda H, Urasawa N, Koyama J, et al. Follow-up of three patients with a large in-frame deletion of exons 45–55 in the Duchenne muscular dystrophy (DMD) gene. J Clin Neurosci. 2008;15:757–63. doi: 10.1016/j.jocn.2006.12.012.
30. Taglia A, Petillo R, D'Ambrosio P, Picillo E, Torella A, Orsini C, et al. Clinical features of patients with dystrophinopathy sharing the 45–55 exon deletion of DMD gene. Acta Myol. 2015;34:9–13.
31. Mendell JR, Campbell K, Rodino-Klapac L, Sahenk Z, Shilling C, Lewis S, et al. Dystrophin immunity in Duchenne's muscular dystrophy. N Engl J Med. 2010;363:1429–37. doi: 10.1056/NEJMoa1000228.
32. Echigoya Y, Yokota T. Skipping multiple exons of dystrophin transcripts using cocktail antisense oligonucleotides. Nucleic Acid Ther. 2014;24:57–68. doi: 10.1089/nat.2013.0451.
33. Naldini L. Gene therapy returns to centre stage. Nature. 2015;526:351–60. doi: 10.1038/nature15818. [PubMed] [Cross Ref]
34. Chen X, Gonçalves MA. Engineered viruses as genome editing devices. Mol Ther. 2016;24:447–57.
35. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol. 2007;25:1298–306. doi: 10.1038/nbt1353.
36. Holkers M, Maggio I, Liu J, Janssen JM, Miselli F, Mussolino C, et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Res. 2013;41:e63. doi: 10.1093/nar/gks1446.
37. Maggio I, Holkers M, Liu J, Janssen JM, Chen X, Gonçalves MA. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Sci Rep. 2014;4:5105. doi: 10.1038/srep05105.
38. Xu L, Park KH, Zhao L, Xu J, El Refaey M, Gao Y, et al. CRISPR-mediated genome editing restores dystrophin expression and function in mdx mice. Mol Ther. 2016;24:564–9. doi: 10.1038/mt.2015.192.
39. Long C, Amoasii L, Mireault AA, McAnally JR, Li H, Sanchez-Ortiz E, et al. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science. 2016;351:400–3. doi: 10.1126/science.aad5725.
40. Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Castellanos Rivera RM, et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science. 2016;351:403–7. doi: 10.1126/science.aad5143.
41. Tabebordbar M, Zhu K, Cheng JK, Chew WL, Widrick JJ, Yan WX, et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science. 2016;351:407–11. doi: 10.1126/science.aad5177.
42. Maggio I, Stefanucci L, Janssen JM, Liu J, Chen X, Mouly V, et al. Selection-free gene repair after adenoviral vector transduction of designer nucleases: rescue of dystrophin synthesis in DMD muscle cell populations. Nucleic Acids Res. 2016;44:1449–70. doi: 10.1093/nar/gkv1540.

43. McGreevy JW, Hakim CH, McIntosh MA, Duan D. Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech. 2015;8:195–213. doi: 10.1242/dmm.018424.
44. Gonçalves MA, de Vries AA. Adenovirus: from foe to friend. Rev Med Virol. 2006;16:167–86. doi: 10.1002/rmv.494.
45. Alonso-Padilla J, Papp T, Kajan GL, Benko M, Havenga M, Lemckert A, et al. Development of novel adenoviral vectors to overcome challenges observed with HAdV-5-based constructs. Mol Ther. 2016;24:6–16.
46. Hendrickx R, Stichling N, Koelen J, Kuryk L, Lipiec A, Greber UF. Innate immunity to adenovirus. Hum Gene Ther. 2014;25:265–84. doi: 10.1089/hum.2014.001.
47. Majhen D, Calderon H, Chandra N, Fajardo CA, Rajan A, Alemany R, et al. Adenovirus-based vaccines for fighting infectious diseases and cancer: progress in the field. Hum Gene Ther. 2014;25:301–17. doi: 10.1089/hum.2013.235.
48. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, Hong JS, et al. Isolation of a common receptor for coxsackie B viruses and adenoviruses 2 and 5. Science. 1997;275:1320–3. doi: 10.1126/science.275.5304.1320.
49. Waddington SN, McVey JH, Bhella D, Parker AL, Barker K, Atoda H, et al. Adenovirus serotype 5 hexon mediates liver gene transfer. Cell. 2008;132:397–409. doi: 10.1016/j.cell.2008.01.016.
50. Wang Z, Zhu T, Qiao C, Zhou L, Wang B, Zhang J, et al. Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. Nat Biotechnol. 2005;23:321–8. doi: 10.1038/nbt1073.
51. Inagaki K, Fuess S, Storm TA, Gibson GA, McTiernan CF, Kay MA, et al. Robust systemic transduction with AAV9 vectors in mice: efficient global cardiac gene transfer superior to that of AAV8. Mol Ther. 2006;14:45–53. doi: 10.1016/j.ymthe.2006.03.014.
52. Ferrari FK, Samulski T, Shenk T, Samulski RJ. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol. 1996;70:3227–34.
53. Mauro A. Satellite cell of skeletal muscle fibers. J Biophys Biochem Cytol. 1961;9:493–5. doi: 10.1083/jcb.9.2.493.
54. Xu X, Wilschut KJ, Kouklis G, Tian H, Hesse R, Garland C, et al. Human satellite cell transplantation and regeneration from diverse skeletal muscles. Stem Cell Rep. 2015;5:419–34. doi: 10.1016/j.stemcr.2015.07.016.
55. Yin H, Price F, Rudnicki MA. Satellite cells and the muscle stem cell niche. Physiol Rev. 2013;93:23–67. doi: 10.1152/physrev.00043.2011.
56. Dumont NA, Wang YX, von Maltzahn J, Pasut A, Bentzinger CF, Brun CE, et al. Dystrophin expression in muscle stem cells regulates their polarity and asymmetric division. Nat Med. 2015;21:1455–63. doi: 10.1038/nm.3990.
57. Yamashita K, Suzuki A, Satoh Y, Ide M, Amano Y, Masuda-Hirata M, et al. The 8th and 9th tandem spectrin-like repeats of utrophin cooperatively form a functional unit to interact with polarity-regulating kinase PAR-1b. Biochem Biophys Res Commun. 2010;391:812–7. doi: 10.1016/j.bbrc.2009.11.144.
58. Kotterman MA, Schaffer DV. Engineering adeno-associated viruses for clinical gene therapy. Nat Rev Genet. 2014;15:445–51. doi: 10.1038/nrg3742.
59. Manno CS, Arruda VR, Pierce GF, Glader B, Ragni M, Rasko J, et al. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med. 2006;12:342–7. doi: 10.1038/nm1358.
60. Mingozzi F, Meulenberg JJ, Hui DJ, Basner-Tschakarjan E, Hasbrouck NC, Edmonson SA, et al. AAV-1-mediated gene transfer to skeletal muscle in humans results in dose-dependent activation of capsid-specific T cells. Blood. 2009;114:2077–86. doi: 10.1182/blood-2008-07-167510.

61. Brantly ML, Chulay JD, Wang L, Mueller C, Humphries M, Spencer LT, et al. Sustained transgene expression despite T lymphocyte responses in a clinical trial of rAAV1-AAT gene therapy. Proc Natl Acad Sci U S A. 2009;106:16363–8. doi: 10.1073/pnas.0904514106.
62. Mendell JR, Rodino-Klapac LR, Rosales-Quintero X, Kota J, Coley BD, Galloway G, et al. Limb-girdle muscular dystrophy type 2D gene therapy restores α-sarcoglycan and associated proteins. Ann Neurol. 2009;66:290–7. doi: 10.1002/ana.21732.
63. Wang Z, Storb R, Halbert CL, Banks GB, Butts TM, Finn EE, et al. Successful regional delivery and long-term expression of a dystrophin gene in canine muscular dystrophy: a preclinical model for human therapies. Mol Ther. 2012;20:1501–7. doi: 10.1038/mt.2012.111.
64. Shin JH, Pan X, Hakim CH, Yang HT, Yue Y, Zhang K, Terjung RL, Duan D. Microdystrophin ameliorates muscular dystrophy in the canine model of duchenne muscular dystrophy. Mol Ther. 2013;21:750–7. doi: 10.1038/mt.2012.283.
65. Kornegay JN, Li J, Bogan JR, Bogan DJ, Chen C, Zheng H, et al. Widespread muscle expression of an AAV9 human mini-dystrophin vector after intravenous injection in neonatal dystrophin-deficient dogs. Mol Ther. 2010;18:1501–8. doi: 10.1038/mt.2010.94.
66. Bengtsson NE, Seto JT, Hall JK, Chamberlain JS, Odom GL. Progress and prospects of gene therapy clinical trials for the muscular dystrophies. Hum Mol Genet. 2016;25(R1):R9–R17. doi: 10.1093/hmg/ddv420.
67. Mendell JR, Rodino-Klapac LR, Rosales XQ, Coley BD, Galloway G, Lewis S, et al. Sustained alpha-sarcoglycan gene expression after gene transfer in limb-girdle muscular dystrophy, type 2D. Ann Neurol. 2010;68:629–38. doi: 10.1002/ana.22251.
68. Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, et al. An essential receptor for adeno-associated virus infection. Nature. 2016;530:108–12. doi: 10.1038/nature16465.
69. Bentzinger CF, Wang YX, von Maltzahn J, Rudnicki MA. The emerging biology of muscle stem cells: implications for cell-based therapies. Bioessays. 2013;35:231–41. doi: 10.1002/bies.201200063.
70. Briggs D, Morgan JE. Recent progress in satellite cell/myoblast engraftment—relevance for therapy. FEBS J. 2013;280:4281–93. doi: 10.1111/febs.12273.
71. Negroni E, Gidaro T, Bigot A, Butler-Browne GS, Mouly V, Trollet C. Invited review: stem cells and muscle diseases: advances in cell therapy strategies. Neuropathol Appl Neurobiol. 2015;41:270–87. doi: 10.1111/nan.12198.
72. Negroni E, Vallese D, Vilquin JT, Butler-Browne G, Mouly V, Trollet C. Current advances in cell therapy strategies for muscular dystrophies. Expert Opin Biol Ther. 2011;11:157–76. doi: 10.1517/14712598.2011.542748.
73. Skuk D, Tremblay JP. Intramuscular cell transplantation as a potential treatment of myopathies: clinical and preclinical relevant data. Expert Opin Biol Ther. 2011;11:359–74. doi: 10.1517/14712598.2011.548800.
74. Huard J, Verreault S, Roy R, Tremblay M, Tremblay JP. High efficiency of muscle regeneration after human myoblast clone transplantation in SCID mice. J Clin Invest. 1994;93:586–99. doi: 10.1172/JCI117011.
75. Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, et al. Disease-specific induced pluripotent stem cells. Cell. 2008;134:877–86. doi: 10.1016/j.cell.2008.07.041.
76. De Bari C, Dell'Accio F, Vandenabeele F, Vermeesch JR, Raymackers JM, Luyten FP. Skeletal muscle repair by adult human mesenchymal stem cells from synovial membrane. J Cell Biol. 2003;160:909–18. doi: 10.1083/jcb.200212064.
77. de la Garza-Rodea AS, van der Velde I, Boersma H, Gonçalves MA, van Bekkum DW, de Vries AA, et al. Long-term contribution of human bone marrow mesenchymal stromal cells to skeletal muscle regeneration in mice. Cell Transplant. 2011;20:217–31. doi: 10.3727/096368910X522117.
78. Dellavalle A, Sampaolesi M, Tonlorenzi R, Tagliafico E, Sacchetti B, Perani L, et al. Peri-

cytes of human skeletal muscle are myogenic precursors distinct from satellite cells. Nat Cell Biol. 2007;9:255–67. doi: 10.1038/ncb1542.
79. Torrente Y, Belicchi M, Sampaolesi M, Pisati F, Meregalli M, D'Antona G, et al. Human circulating AC133+ stem cells restore dystrophin expression and ameliorate function in dystrophic skeletal muscle. J Clin Invest. 2004;114:182–95. doi: 10.1172/JCI20325.
80. Sampaolesi M, Torrente Y, Innocenzi A, Tonlorenzi R, D'Antona G, Pellegrino MA, et al. Cell therapy of α-sarcoglycan null dystrophic mice through intra-arterial delivery of mesoangioblasts. Science. 2003;301:487–92. doi: 10.1126/science.1082254.
81. Tedesco FS, Hoshiya H, D'Antona G, Gerli MF, Messina G, Antonini S, et al. Stem cell-mediated transfer of a human artificial chromosome ameliorates muscular dystrophy. Sci Transl Med. 2011;3:96ra78. doi: 10.1126/scitranslmed.3002342.
82. Benchaouir R, Meregalli M, Farini A, D'Antona G, Belicchi M, Goyenvalle A, et al. Restoration of human dystrophin following transplantation of exon-skipping-engineered DMD patient stem cells into dystrophic mice. Cell Stem Cell. 2007;1:646–57. doi: 10.1016/j.stem.2007.09.016.
83. Torrente Y, Belicchi M, Marchesi C, D'Antona G, Cogiamanian F, Pisati F, et al. Autologous transplantation of muscle-derived CD133+ stem cells in Duchenne muscle patients. Cell Transplant. 2007;16:563–77. doi: 10.3727/000000007783465064.
84. Cossu G, Previtali SC, Napolitano S, Cicalese MP, Tedesco FS, Nicastro F, et al. Intra-arterial transplantation of HLA-matched donor mesoangioblasts in Duchenne muscular dystrophy. EMBO Mol Med. 2015;7:1513–28. doi: 10.15252/emmm.201505636.
85. Skuk D, Tremblay JP. Confirmation of donor-derived dystrophin in a Duchenne muscular dystrophy patient allotransplanted with normal myoblasts. Muscle Nerve. 2016
86. Law PK, Bertorini TE, Goodwin TG, Chen M, Fang QW, Li HJ, et al. Dystrophin production induced by myoblast transfer therapy in Duchenne muscular dystrophy. Lancet. 1990;336:114–5. doi: 10.1016/0140-6736(90)91628-N.
87. Montarras D, Morgan J, Collins C, Relaix F, Zaffran S, Cumano A, et al. Direct isolation of satellite cells for skeletal muscle regeneration. Science. 2005;309:2064–7. doi: 10.1126/science.1114758.
88. Gilbert PM, Havenstrite KL, Magnusson KE, Sacco A, Leonardi NA, Kraft P, et al. Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. Science. 2010;329:1078–81. doi: 10.1126/science.1191035.
89. Balsam LB, Wagers AJ, Christensen JL, Kofidis T, Weissman IL, Robbins RC. Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium. Nature. 2004;428:668–73. doi: 10.1038/nature02460.
90. Murry CE, Soonpaa MH, Reinecke H, Nakajima H, Nakajima HO, Rubart M, et al. Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts. Nature. 2004;428:664–8. doi: 10.1038/nature02446.

Summary and **Concluding** Marks

Clinical treatments are transitioning from traditional "one-fits-all" chemical drugs, to personalized or precision medicine based on the genetic information of each patient. Either in biomedicine as a whole or in precision medicine exclusively, genome editing technologies are going to play an ever-increasing role by helping decipher the function of genomic sequences and by providing the tools for permanently correcting the underlying cause of genetic disorders - the ideal situation one can envision in the future. Related to this, engineered RNA-guided CRISPR-Cas9 nucleases are becoming one of the main driving forces to bring these ideas into reality. However, there are still crucial pitfalls associated with CRISPR-Cas9 nucleases that need to be tackled before their implementation as agents in standard-of-care therapies. These bottlenecks include ineffective cellular delivery of the genome editing reagents, their off-target DNA cleaving activities and the inefficient and/or imprecise the exogenous DNA incorporation into the human genome. To tackle these issues, researchers are developing improved delivery agents, scaling-up their production and, critically, enhancing the specificity of double-stranded DNA break (DSB) formation. Finally, researchers are also dissecting the cellular factors that can affect genome editing outcomes, notably among these, the chromatin status of target chromosomal sequences. It is under this context that, this thesis has investigated, on one end, a DSB-free genome editing strategy coined in trans paired nicking, and on the other, the impact of higher-order chromatin conformations on the activity of genome editing reagents and the choice of DNA repair pathway, i.e., non-homologous end joining (NHEJ) versus homology-directed repair (HDR).

Chapter 1 reviews the expanding genome-editing tool armamentarium and associated DNA modification strategies, with an emphasis on programmable nucleases based on CRISPR systems. **Chapter 2** demonstrates that in trans paired nicking genome editing can precisely incorporate small and large DNA segments at different loci in human cells, including the *AAVS1* "safe harbor" locus in pluripotent stem cells. Different HDR genome editing strategies involving either DSBs or single-stranded DNA breaks (SSBs), were tested in parallel and assessed in terms of their relative efficiency, specificity and accuracy. Conventional genome editing based on DSB formation readily yielded mutation byproducts presumably derived from NHEJ; genome editing based on SSB formation was, in turn, inefficient. Critically, in trans paired nicking, resorting to the generation of SSBs at chromosomal and plasmid donor DNA using CRISPR-Cas9 "nickases", led to efficient and seamless genome editing without attendant detection of NHEJ-derived mutations at target alleles.

Consistent with the initial homologous recombination model proposed by Holliday in 1964, it is possible that the coordinated generation of SSBs at target and donor DNA substrates facilitates key rate-limiting HDR steps, e.g., reciprocal single-stranded DNA invasion by both interacting partners. In this regard, further investigations are warranted to tease apart the players and mechanisms governing nick-induced HDR

pathways through, for instance, unbiased library screens or knocking-out candidate DNA repair genes. From a practical point of view, it will be interesting to assess the potential of in trans paired nicking for endogenous gene tagging, and editing multi-copy genes or genes with overlapping sequences. Next to this, it will be worth determining whether in trans paired nicking is compatible with Cas9 orthologs and high-specificity Cas9 variants to further improve its versatility and accuracy, respectively. Admittedly, a possible drawback of the in trans paired nicking strategy might be its unsuitability for the genetic modification of quiescent cells due to their lack of an active, canonical, HDR machinery. Hence, alternative strategies that can achieve efficient and non-mutagenic genome editing in post-mitotic cells are in demand and should thus be enthusiastically pursued.

Chapter 3 probes, in a quantitative manner, the impact of alternate chromatin states on the performance of genome editing tools, in particular, programmable nucleases based on *S. pyogenes* CRISPR-Cas9 nucleases and *Xanthomonas sp*. transcription activator-like effector (TALE) proteins. For these experiments, complementary gain-of-function and loss-of-function cellular models were set-up to track targeted gene knockout levels by different CRISPR-Cas9 nucleases and TALE nucleases (TALENs) at euchromatin versus heterochromatin. In these human-cell systems, reporter target alleles transit from compact heterochromatin to relaxed euchromatin through the doxycycline-dependent release of a tTR-KRAB fusion protein whose effector domain Krüppel-associated box (KRAB) recruits, amongst other heterochromatin-assembling factors, KAP-1 and HP-1. The data generated in these reporter systems demonstrates that TALENs and CRISPR-Cas9 nucleases are both significantly hindered by KRAB-induced heterochromatin in living cells with the former more so than the latter. This finding is intriguing in view of the fact that, in contrast to TALE proteins, *S. pyogenes* CRISPR-Cas9 nucleases do not evolve to assess and cleave genomic DNA in the nuclei of eukaryotic cells. It is of note, however, **Chapter 4** indicate that *S. aureus* CRISPR-Cas9 nucleases are significantly more affected by closed chromatin than their *S. pyogenes* counterparts. Whether this differential "chromatin barrier" results from these nucleases having different protospacer adjacent motifs, conformational energy thresholds for ATP-independent helicase activation, or other downstream processes, warrants further investigation.

As aforementioned, a preeminent shortcoming of RNA-guided nucleases is their off-target activity. To address this issue head-on, researchers are developing new reagents and strategies to enhance the specificity of genome editing protocols. For instance, target site discrimination by RNA-guided nucleases can be improved through rationally engineered high-specificity Cas9 variants, offset RNA-guided "nickase" pairs, and truncated gRNAs (Tru-gRNAs) whose shorter spacer sequences often bias DNA binding towards fully complementary sequences. **Chapter 3** and **Chapter 4** demonstrate that, albeit to different extents, all these high-specificity

RNA-guided nuclease platforms suffer from having their target sequences embedded in KRAB-induced heterochromatin. In fact, with the exception of Tru-gRNAs, heterochromatin seems to impinge a higher activity barrier to these high-specificity platforms than to conventional RNA-guided nucleases. Future studies should seek to determine whether, not only on-target, but also off-target profiles of programmable nucleases change due to different epigenomes, such as those characteristic of specific cell types and cell differentiation stages.

Using the same tTR-KRAB reporter cells, **Chapter 5** studys the impact of alternate chromatin conformations on the relationship between NHEJ versus HDR in genome editing. In these experiments, donor DNA templates were provided in the form of single-stranded oligodeoxyribonucleotides, standard plasmids or Integrase-defective lentiviral vector (IDLV) genomes. In contrast to the observation that heterochromatin partially impairs NHEJ-mediated gene editing (**Chapters 3** and **4**), the absolute levels of HDR-mediated gene editing are similar at euchromatin and heterochromatin. Hence, the balance between NHEJ and HDR shifts towards HDR when target sequences transit from euchromatin to heterochromatin. As corollary, when compared to euchromatic sites, at heterochromatic sites, there are fewer mutagenic NHEJ-derived indels for each HDR-derived gene editing event.

Although the "downstream" chromatin transition in tTR-KRAB reporter cells entails recruiting endogenous chromatin-remodeling complexes and establishing "physiological" KRAB-dependent chromatin states, the "upstream" epigenetic regulation of target sequences is surrogate in character. Thus, it should be insightful to complement the data from tTR-KRAB-regulated reporter alleles with those from endogenous loci harboring different epigenetic marks and chromatin organizations. This line of inquiry should profit from the expanding set of epigenetic-remodeling tools which, in a targeted manner, can modulate DNA methylation and/or post-translationally alter nucleosome compositions, e.g., histone acetylation/deacetylation. Collectively, the data presented in **Chapters 3**, **4** and **5** raise the possibility for directing genome editing outcomes by controlling the higher-order epigenetic states of target sequences. In particular, "opening-up" a heterochromatic target site might increase gene knockout efficiencies, while transiently "closing" a euchromatic target site might reduce NHEJ-derived indels upon HDR-based genome editing. Such ancillary approaches might become feasible with the aid of the aforementioned toolbox consisting of epigenetic-remodeling programmable factors. Of note, besides higher-order chromatin structures, recent *in vitro* studies have demonstrated that nucleosome positioning can, per se, influence target DNA cleavage by CRISPR-Cas9 nucleases. Interestingly, modulating the dynamics of nucleosome positioning *in vitro* through ATP-dependent chromatin remodelers greatly enhanced target DNA cleavage by CRISPR-Cas9 nucleases.

Chapter 6 provides an overview of the main characteristics, pros and cons of commonly used viral vector systems that are being converted into delivery vehicles for genome editing tools. In particular, IDLVs, adeno-associated viral vectors and adenoviral vectors. As a case-in-point, **Chapter 7** covers the emerging investigations using viral vectors for the *ex vivo* or *in vivo* correction of defective dystrophin-encoding gene causing the lethal muscle-wasting X-linked disorder Duchenne muscular dystrophy (DMD). Regarding the *in vivo* delivery of programmable nucleases, it is important to consider the potential immunogenicity of genome editing reagents and delivery vehicle components, e.g., bacterial Cas9 proteins and viral vector capsids, respectively. In fact, recent experiments have identified pre-existing humoral and cellular immune responses to *S. pyogenes* and *S. aureus* Cas9 nucleases in the human population. These aspects are especially relevant in *in vivo* settings where genome editing components directly encounter the recipient's immune system.

In conclusion, the interplay between different donor DNA structures and target chromatin environments modulates the outcomes of genome editing in terms of their efficiency, specificity and accuracy. Hence, the findings of this thesis provide a new path for improving these three key parameters that underlie robust and seamless genetic modification of human cells.

Addendum

Nederlandse samenvatting

List of publications

Curriculum Vitae

Acknowledgments

Nederlandse samenvatting

Klinische behandelingen gaan over van een traditionele aanpak, zoals het toedienen van "one-fits-all" medicijnen, naar gepersonaliseerde- of precisiegeneeskunde op basis van de genetische informatie van elke patiënt. Of het nu in de context van de medische biologie in zijn geheel is of de precisiegeneeskunde in het bijzonder, het is redelijk om aan te nemen dat genoom-editing technieken een steeds grotere rol gaan spelen binnen deze gebieden door de functie van genoomsequenties te helpen ontcijferen en door de hulpmiddelen aan te bieden voor het permanent corrigeren van de onderliggende oorzaak van genetische aandoeningen - de ideale situatie die men zich voor de toekomst kan voorstellen. RNA-geleide CRISPR-Cas9-nucleasen zijn een van de belangrijkste drijvende krachten aan het worden om deze ideeën tot werkelijkheid te maken. Er zijn echter nog steeds cruciale valkuilen die moeten worden aangepakt voordat CRISPR-Cas9-nucleasen geïmplementeerd kunnen worden in therapieën die voldoen aan de eisen voor zorg. Deze knelpunten omvatten inefficiënte cellulaire afgifte van de genoom-editing reagentia, hun "off-target" DNA knipactiviteiten en de inefficiënte en / of onnauwkeurige opname van het exogene DNA in het menselijk genoom. Met het oog op het oplossen van deze problemen ontwikkelen onderzoekers verbeterde methoden voor de cellulaire afgifte van de reagentia, voeren ze het productieniveau op en, van groot belang, verbeteren ze de specificiteit van de dubbelstrengs DNA-breuk (DSB) vorming. Om te helpen bij de keuze van programmeerbare nucleaseplatforms en om hun prestaties te verbeteren, beginnen onderzoekers ook de cellulaire factoren te ontleden die de uitkomst van genoom-editing kunnen beïnvloeden, met in het bijzonder de chromatine-status van chromosomale doelwitsequenties.

Het is in deze generieke context dat ik in dit proefschrift mijn onderzoek naar een DSB-vrije genoom-editingstrategie genaamd "in trans paired nicking" beschrijf en ook de invloed van hogere-orde chromatine-conformaties op de activiteit van genoom-editingreagentia en keuze van de DNA-herstelroute, d.w.z. niet-homologe eindverbinding (NHEJ) versus homologie gerichte reparatie (HDR), aan de orde komt. Hoofdstuk 1 biedt een overzicht van het uitbreidende assortiment aan genoom-editing reagentia en de bijbehorende DNA-modificatiestrategieën, met de nadruk op programmeerbare nucleasen op basis van CRISPR-systemen. Hoofdstuk 2 laat zien dat "in trans paired nicking" genoom-editing kan resulteren in de precieze incorporatie van kleine en grote DNA-segmenten op verschillende loci in menselijke cellen, inclusief het *AAVS1* "safe harbor" locus in pluripotente stamcellen. Verschillende genoom-editingstrategieën waarbij de HDR-route werd geactiveerd door DSB's of enkelstrengs DNA-breuken (SSB's), werden parallel getest en beoordeeld op hun relatieve efficiëntie, specificiteit en nauwkeurigheid. Conventionele genoom-editing op basis van DSB-vorming op de doellocatie leverde frequent mutatiebijproducten op, vermoedelijk als gevolg van NHEJ; genoom-editing op basis van SSB-formatie

op de doellocatie was op zijn beurt inefficiënt. Bij "in trans paired nicking", waarbij gebruik wordt gemaakt van de generatie van SSB's op chromosomaal en plasmide donor DNA met behulp van CRISPR-Cas9 "nickases", leidde tot een efficiënte en naadloze bewerking van het genoom zonder de aanwezigheid van NHEJ-afgeleide mutaties bij doelallelen.

In overeenstemming met het initiële homologe recombinatiemodel dat door Holliday in 1964 werd voorgesteld, is het mogelijk dat de gecoördineerde generatie van SSB's op doel- en donor- DNA- substraten de uitvoeren van HDR-stappen vergemakkelijkt, bijv. reciproke enkelstrengs DNA-invasie door beide samenwerkende partners. Verdere onderzoeken om de spelers en mechanismen die betrokken zijn bij nick-geïnduceerde HDR-routes te achterhalen zijn nodig, bijvoorbeeld door onbevooroordeelde bibliotheek screening of gerichte knock-out van kandidaat-genen die betrokken zijn bij het DNA herstel mechanisme. Het beoordelen van het potentieel van "in trans-paired nicking" voor het labellen van endogene eiwitten, met als doel het live volgen van levende cellen, en het bewerken van multi-kopie genen of genen met overlappende sequenties is vanuit een praktisch oogpunt interessant. Daarnaast is het de moeite waard om te bepalen of "in trans-paired nicking" compatibel is met Cas9-orthologen en zeer specifieke Cas9-varianten om, respectievelijk, de veelzijdigheid uit te breiden en de nauwkeurigheid te verbeteren. Toegegeven, een mogelijk nadeel van de "in trans-paired nicking" strategie zou de ongeschiktheid voor genetische modificatie van rustende cellen kunnen zijn, vanwege het ontbreken van de actieve, gebruikelijke, HDR-machinerie. Vandaar dat alternatieve strategieën die efficiënte en niet-mutagene genoom-editing in post-mitotische cellen kunnen bewerkstelligen gevraagd zijn en dus enthousiast moeten worden nagestreefd. Hoofdstuk 3 beschrijft experimenten die zijn ontworpen om, op een kwantitatieve manier, de impact te bepalen van chromatine status op de prestaties van genoom-editingreagentia, in het bijzonder van programmeerbare nucleasen op basis van componenten van *S. pyogenes* CRISPR-Cas9 nucleasen en *Xanthomonas sp.* transcriptie activator-like effector (TALE) eiwitten. Voor deze experimenten werden complementaire "gain-of-function" en "loss-of-function" cellulaire modellen gegenereerd om de doelgen-knock-out levels, tot stand gekomen door de werking van verschillende CRISPR-Cas9-nucleasen en TALE-nucleasen (TALENs) bij euchromatin versus heterochromatine, te volgen.

In deze menselijke cel-systemen gaan reporter doelwitallelen over van compact heterochromatine naar ontspannen euchromatine via de doxycycline afhankelijke werking van een tTR-KRAB fusie-eiwit waarvan het effectordomein, Krüppel-geassocieerde box (KRAB), onder andere de heterochromatine-assemblerende factoren KAP-1 en HP-1 rekruteert. De met deze reportersystemen gegenereerde gegevens, gepresenteerd in hoofdstuk 3, tonen aan dat TALEN's en CRISPR-Cas9-nucleasen beide aanzienlijk worden belemmerd door KRAB-geïnduceerde heterochromatine

in levende cellen, de eerste meer dan de laatstgenoemde. Deze bevinding is intrigerend met het oog op het feit dat, in tegenstelling tot TALE-eiwitten, *S. pyogenes* CRISPR-Cas9-nucleasen niet zijn geëvolueerd om genomisch DNA in de kernen van eukaryote cellen te benaderen en te knippen. Het moet echter worden opgemerkt dat RNA-geleide nucleasen van andere prokaryotische CRISPR-systemen meer worden belemmerd door de eukaryote context van doelsequenties, in het bijzonder bij heterochromatische gebieden, dan die afgeleid van *S. pyogenes* CRISPR-Cas9-componenten. De in hoofdstuk 4 vermelde experimentele gegevens, verkregen bij gebruik van dezelfde cellulaire modellen als in hoofdstuk 3, tonen inderdaad aan dat *S. aureus* CRISPR-Cas9-nucleasen significant meer worden beïnvloed door gesloten chromatine dan hun S. pyogenes tegenhangers. Of deze differentiële "chromatinebarrière" het gevolg is van de verschillende "protospacer adjacent motifs" van deze nucleasen, de conformationele energiedrempels voor ATP-onafhankelijke helicase-activering of andere downstream processen die leiden tot het knippen van DNA verdient verder onderzoek. Zoals reeds eerder vermeld, is een bekende tekortkoming van RNA-geleide nucleasen hun off-target activiteit. Om dit probleem rechtstreeks aan te pakken, zijn onderzoekers actief bezig met het ontwikkelen van nieuwe reagentia en het onderzoeken van nieuwe strategieën om de specificiteit van genoom-editing protocollen te verbeteren.

Deze pogingen resulteerden in een verbetering van doellocatie herkenning door RNA-geleide nucleasen via het rationeel construeren van hoog-specifieke Cas9-varianten, het gebruik van offset-RNA-geleide "nickase"-paren en het ontwerpen van ingekorte gRNA's (Tru-gRNA's) waarvan de kortere DNA-hybridiserende spacersequenties vaak productieve DNA-binding van volledig complementaire doelsequenties bevorderen. De in hoofdstuk 3 en hoofdstuk 4 gepresenteerde resultaten laten zien dat al deze hoog-specificiteit RNA-geleide nucleaseplatforms lijden, alhoewel in verschillende mate, wanneer hun doelwitsequenties zijn ingebed in door KRAB geïnduceerd heterochromatine. Heterochromatine lijkt een grotere impact te hebben op platforms met hoge specificiteit dan op conventionele RNA-geleide nucleasen, met uitzondering van de Tru-gRNA's. Toekomstige studies moeten proberen vast te stellen of niet alleen on-target maar ook off-target profielen van programmeerbare nucleasen veranderen door verschillende epigenomen, zoals die die kenmerkend zijn voor specifieke celtypen en celdifferentiatiestadia.

List of publications

1. Chen X, Liu J, Janssen JM, Gonçalves MA. The chromatin structure differentially impacts high-specificity CRISPR-Cas9 nuclease strategies. Molecular Therapy-Nucleic Acids 8:558–563 (2017).

2. Chen X, Janssen JM, Liu J, Maggio I, 't Jong A, Mikkers H & Gonçalves MA. In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. Nature Communications 8:657 (2017).

3. Chen X, Rinsma M, Janssen JM, Liu J, Maggio I, Gonçalves MA. Probing the impact of chromatin conformation on genome editing tools. Nucleic Acids Research 44:6482-6492 (2016).

4. Chen X, Gonçalves MA. Engineered viruses as genome editing devices. Molecular Therapy 24:447-457 (2016).

5. Maggio I*, Chen X*, Gonçalves MA. The emerging role of viral vectors as vehicles for DMD gene editing. Genome Medicine 8:59 (2016). (*co-first author)

6. Maggio I, Stefanucci L, Janssen JM, Liu J, Chen X, Mouly V, Gonçalves MA. Selection-free gene repair after adenoviral vector transduction of designer nucleases: rescue of dystrophin synthesis in DMD muscle cell populations. Nucleic Acids Research 44:1449-1470 (2016).

7. Maggio I, Holkers M, Liu J, Janssen JM, Chen X, Gonçalves MA. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Scientific Reports 4:5105 (2014)

Curriculum Vitae

Xiaoyu Chen was born on September 6th, 1990 in the city of Xingtai, Hebei Province, China. During 2006-2013 she was enrolled in a 7-year bachelor-master combined program of clinical medicine in Hebei Medical University, Shijiazhuang, China. From 2006 to 2008, she studied general college courses in Nankai University, Tianjin, China. She received her BSc degree in medicine in 2011 and MSc degree in neurology in 2013. From 2011 to 2013, she worked in the laboratory of Professor Chunyan Li to study the neurodegenerative disease ALS using the *SOD1*^{G93A} mouse model. During her studies in Hebei Medical University, she received a national undergraduate scholarship in 2008 and a national graduate scholarship in 2011 from the Ministry of Education of China.

In the last year of her master study, she developed a great interest in gene therapy, whose goal is to target the root cause of diseases. Instead of becoming a medical doctor, after her graduation, she applied for a Ph.D. scholarship from the China Scholarship Council to pursue a research career in the Netherlands. In 2013, Xiaoyu started her Ph.D. studies in the group of Prof. Dr. Rob C. Hoeben under the supervision of Dr. Manuel A.F.V. Gonçalves in the Department of Cell and Chemical Biology at the Leiden University Medical Center (LUMC).

Throughout her Ph.D. studies, she focused on investigating high-fidelity genome engineering strategies, especially those based on Cas9 "nickases" and target site-containing donor DNA molecules. Next to this research, she has also probed the impact of higher-order chromatin on the performance of gene editing tools, and on the balance of different DNA repair pathways. These studies were presented in this thesis. Currently, Xiaoyu Chen is looking forward to translate genome engineering technology from the bench to bed side.

Acknowledgments

This is it, an exciting adventure of Ph.D. I thought it might be about survival, but turns out to be epic. Surrounded by so many amazing people, glancing back, I know that, the fortune always smiles upon me.

I would like to express my heartfelt gratitude to my promotor Prof. Dr. Rob C. Hoeben who showed me the essence of leadership: open-mind, generosity, and rationality. I want to express my deepest gratitude to my supervisor Dr. Manuel A.F.V. Gonçalves for unveiling me the beauty of science, including its sweets and bitterness. He guided me with patience and earnest instruction, and taught me how to make sense out of mess and how to make a win out of loss.

Special tribute go to my paranymphs Igno and Jin, who always encourage me whenever I am drained out of energy and cheer me up on every single milestone I made. I thank Kim for her tremendous help in assisting my projects and organizing the lab. I also thank the people that I have worked with for their team spirits and camaraderie: Francesca, Hidde, Ivan, Luca, Marcela, Marrit, Marteen, Qian and Sara. My thanks also go to the colleagues in LUMC–Amiet, Anke, Anna, Arnaud, Diana, Erik, Gelilla, Harald, Henrique, Iris, Jan, Jiang, Lussi, Maaike, Marjolein, Martijn, Peter, Ron Wolterbeek, Sijia, Steve, Thilo, Timo, Tobias, Vera, Willem, Yongsheng, Yvonne–with whom I shared meetings, paper celebrations and birthday cakes. My sincere thanks also go to Prof. Peter ten Dijke, who once told me "Just be yourself" and relieved me from self-doubt.

I also thank my Chinese friends in Leiden–He Nan, Jiali, Yifei, Jianhui, Xinpei, Yanming, Jinxian, Jian Yang, Guangchao, Kaifeng, Channa, Wenbo, Xiaoke, Zhiwei, Puning, Yingguang, Zhang Rui, Zhenyu, Fuyu, Qiuyin. The sharing trips and food have been great solace since we are thousands of miles away from home.

I would also like to express my special thanks to professor Chunyan Li, who planted the seed of scientific research in my mind and whose support stand by me since 2011.

Last, I would like to dedicate this thesis to my parents and brother Yanpei, who are always supportive for my dream, though, equally concerned on my safety.

The journey ends; the quest still goes on.

Now, peace and serenity.