Universiteit Leiden
The Netherlands

Soft genome editing based on CRISPR nickases: it takes one break to tango

Wang, Q.

Citation
Wang, Q. (2024, June 26). *Soft genome editing based on CRISPR nickases: it takes one break to tango*. Retrieved from https://hdl.handle.net/1887/3765882

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3765882

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

Soft genome editing based on CRISPR nickases: it takes one break to tango

Qian Wang

ISBN: 978-94-6496-101-0

Copyright © Qian Wang, Leiden, the Netherlands, 2024. All right reserved. No part of this thesis may be reproduced, stored, translated, or transmitted in any form or by any means now or hereafter, electronic or mechanical without prior written permission of the copyright owner.

Cover: concept and design by Qian Wang
Layout: Qian Wang
Printing: GildePrint

Soft genome editing based on CRISPR nickases: it takes one break to tango

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 26 juni 2024
klokke 13:45 uur

door

Qian Wang
geboren te Lianyungang, Jiangsu, China
in 1992

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

Leden promotiecommissie:
Prof. Dr. C. Brakenbusch (University of Copenhagen, Denmark)
Prof. Dr. F.J.T. Staal
Prof. Dr. S.M. Chuva de Sousa Lopes
Dr. B. Pang
Dr. H.J.G. Snippert (University Medical Center Utrecht, the Netherlands)
Dr. J. Wijnholds

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 supported by the Prinses Beatrix Spierfonds (W.OR11-18, W.OR16-13, W.OR21-01), the Dutch Duchenne Parent Project (17.012), the China Scholarship Council-Leiden University Joint Scholarship Program, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 765269 (IMGENE – Improving Genome Editing Efficiency) and the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska Curie Actions grant agreement No. 101072427 (GetRadi – Gene Therapy of Rare Diseases).

To My Dear Family and Friends

谨以此书献给亲爱的家人及朋友

CONTENTS

General Introduction / ***009***

Chapter 1 Adenoviral vectors meet gene editing: a rising partnership for the genomic engineering of human stem cells and their progeny / ***013***

Chapter 2 Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking / ***039***

Chapter 3 Precise and broad scope genome editing based on high-specificity Cas9 nickases. / ***067***

Chapter 4 Precise homology-directed installation of large genomic edits in human cells with cleaving and nicking high-specificity Cas9 variants / ***099***

Chapter 5 Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery / ***125***

Chapter 6 Selection-free precise gene repair using high-capacity adenovector delivery of advanced prime editing systems rescues dystrophin synthesis in DMD muscle cells / ***145***

Chapter 7 The chromatin context differently impacts prime editors and base editors and controls the fidelity and purity of base editing / ***167***

Conclusions and Final Remarks / ***209***

Addendum *Nederlandese Samenvatting* / ***214***
List of Publications / ***218***
Curriculum Vitae / ***220***
Acknowledgments / ***221***

The manipulation of the human genome with customized genetic information makes it possible to further decipher the basis of biological processes under both physiological and disease states. In the past decades, rapid technological breakthroughs originating from fundamental microbiology research have yielded novel genome engineering tools and principles that greatly facilitate our ability to efficiently modify specific genomic sequences in living cells and organisms. As a corollary, these technologies are also starting to permeate the realm of medicine when applied as a form of "genomic surgery". These genetic therapies aim at tacking the root cause of human pathologies, inherited or acquired, by correcting or modulating the genetic content or expression, respectively, present in target cells, tissues and organs. To this end, delivery vehicles capable of introducing, in an efficient and safe manner, the increasingly sophisticated (epi)genome editing reagents are in demand, especially when considering *in vivo* genetic therapies.

Owing to their robustness, simplicity, and versatility, engineered RNA-guided nucleases (RGNs) built on prokaryotic clustered regularly interspaced short palindromic repeat (CRISPR)–Cas9 systems, consisting of sequence-tailored single guide RNAs (gRNAs) coupled to Cas9 endonucleases, remain amongst the most powerful genome editing tools since their introduction by independent groups in 2013. Commonly, RGN-based genome editing manoeuvres start by the triggering of site-specific chromosomal double-stranded DNA breaks (DSBs) that, upon endogenous DNA repair pathways activation, yield gene knockouts and, in the presence of exogenous donor DNA, gene knock-ins. Generically, the former and latter genome editing outcomes involve the recruitment of non-homologous end joining (NHEJ) and homology-directed repair (HDR) factors, respectively, at the RGN-induced DSBs. Yet, precise genome editing is often hindered due to the multiple-copy character of the vast majority of chromosomal sequences and off-target RGN activities. Additionally, targeted DSBs required for cellular DNA repair activation as well as DSBs resulting from off-target DNA cleavage inevitably produce inaccurate and unpredictable genetic structural variants in the form of small insertions and deletions (indels) and local or genome-wide chromosomal rearrangements, e.g., duplications, large deletions and/or translocations. Moreover, regardless of their specificity, RGNs trigger P53-dependent cell cycle arrest and apoptosis. Indeed, the activation of this DNA damage response (DDR) limits the efficacy of genome editing procedures. This is especially so in the case of DSB-dependent genome editing in regular P53 proficient stem cells that serve as highly relevant substrates for human disease modelling and therapy, e.g., induced pluripotent stem cells (iPSCs) and hematopoietic stem cells, respectively. Equally insidious, DDR activation is known to create selective pressure for the emergence of gain-of-function and loss-of-function gene mutations linked to tumorigenesis.

Hence, the research presented in this thesis is primarily directed to the heightening of the specificity and fidelity of genome editing procedures by investigating and harnessing nicking RGNs and their prime editing derivatives based on prototypic CRISPR–Cas9 systems. Unlike regular Cas9 nucleases, sequence- and strand-specific Cas9 nucleases ("nickases") contain either their RuvC or HNH nuclease domains disabled. Therefore, when compared with intrinsically mutagenic DSBs, single-stranded DNA breaks (SSBs), or nicks, made by such RNA-programmable enzymes, are less disruptive to the genome in that they do not constitute canonical substrates for error-prone DNA repair processes, e.g., NHEJ and microhomology-mediated end joining. Importantly, albeit at low frequencies, SSBs are capable nonetheless of triggering HDR in mammalian cells. This knowledge has laid a foundation for further investigating herein the *in trans* paired nicking (ITPN) concept based on enhancing HDR-mediated genome editing by combining nicking RGNs with nicking-susceptible donor DNA constructs. In addition, the feasibility and utility of deploying adenoviral vector (AdV) technologies for the purpose of prime editing in DSB-sensitive and hard-to-transfect cell types, namely, muscle progenitors, mesenchymal stem cells and iPSCs, is established. Finally, research presented in this thesis further discloses a role for higher-order chromatin conformations on the ultimate efficiency and fidelity of prime editors and base editors, both comprising nicking RGNs fused to effector domains responsible for the installation of specific genomic edits in a DSB- and donor DNA-independent manner.

Chapter 1 serves as an introductory chapter in that it provides a detailed overview about the principles governing the main genome editing strategies and associated effector platforms focusing on those that have entered the clinical trial arena, i.e., zinc-finger nucleases, transcription activator-like effector nucleases and RNA-guided CRISPR-Cas9 nucleases. In addition, **Chapter 1** reviews applications of these genome editing tools and strategies in human stem cells focusing on the use of adenoviral vectors

(AdVs) as delivery vehicles. This chapter further highlights the opportunities offered by high-capacity AdVs (HC-AdVs) in particular for ferrying large donor DNA payloads and DNA-editing fusion constructs, such as those underlying DSB-independent base editing and prime editing processes.

Research presented in **Chapter 2** formally demonstrates that the formation of indels resulting from RGN-induced DSBs at target sites can lead to the loss of fitness by gene-edited cells and reports that simultaneous SSB formation at donor DNA constructs and acceptor chromosomal sequences by nicking RGNs (i.e., ITPN genome editing), can overcome such disruptive genotype-phenotype associations. Moreover, ITPN compared favorably with DNA manipulations involving the exclusive formation of SSBs or DSBs at chromosomal sequences as it yields more frequent and seamless, respectively, HDR-mediated genome editing events in human cells.

A major concern in the genome editing field as a whole, that acquires particular relevance when considering therapeutic gene-editing interventions, regards the activity of RGNs at off-target chromosomal sequences. These unintended off-target activities and associated collateral effects result from the fact that wild-type Cas9 proteins often remain proficient at DNA cleaving even when multiple mismatches and/or bulges exist between gRNA spacer and genomic sequences. This is especially the case when the mismatches and/or bulges locate distally to protospacer adjacent motif (PAM) sites, i.e., sites that constitute the initial engagement points of RGN complexes with DNA. Therefore, the cumulative work described in **Chapter 3** and **Chapter 4** sought to tackle this issue through the assembly, testing and validation of dual and single RGNs with nicking Cas9 variants capable of triggering gene knock-outs and gene knock-ins in an efficient and highly specific manner. In particular, in **Chapter 3**, a systematic assessment of the activities and specificities attained by a representative panel of high-specificity Cas9 nucleases and their corresponding RuvC-disabled Cas9^{D10A} variants, was conducted. Importantly, dual nicking RGNs based on specific Cas9^{D10A} variants were shown to outperform parental dual nicking RGNs and achieve selective cleavage of target sequences with high similarity to off-target sites. Following from these findings, **Chapter 4** further investigates the capability of orthogonal and high-specificity Cas9 proteins in directing gene targeting through homologous recombination (HR) and homology-mediated end joining (HMEJ), and explores the compatibility of the IPTN principle with orthogonal and high-specificity Cas9 nickases.

Prime editing is a recent precision genome editing modality that permits installing any single base-pair substitution and well-defined indels at specific genomic positions requiring to this effect neither DSBs nor donor DNA substrates. However, the large size of prime editing complexes poses substantial production and delivery issues. As discussed in **Chapter 1**, the HC-AdV platform presents a particularly valuable set of features that warrants its exploitation for genome editing purposes, namely (i) efficient transduction of cycling and quiescent cells; (ii) amenability to tropism modifications; (iii) high genetic stability; (iv) strict episomal nature; (v) absence of viral genes, and (vi) vast packaging capacity (i.e., up to 36 kb). In this regard, **Chapter 5** reports the feasibility of tailoring these biological nanoparticles for all-in-one transfer of full-length prime editing components into both transformed and non-transformed cell types. In addition, the positive influence of cellular replication on prime editing activity is disclosed by exploiting the HC-AdV cell cycle independency. Building on these findings, **Chapter 6** outlines the therapeutic potential of HC-AdV delivery of advanced prime editing machineries comprising optimized and multiplexing components. In these HC-AdV-enabled prime editing experiments, Duchenne muscular dystrophy (DMD) was targeted as a disease model. DMD (OMIM #310200) is an X-linked progressive muscle-wasting disorder (incidence: ~1:4700 boys) caused by loss-of-function mutations in the large *DMD* gene (~2.4 Mb) that normally codes for the striated muscle-stabilizing protein dystrophin (427 kDa). Of notice, in-frame *DMD* deletions result in a less acute form of muscular dystrophy, named Becker muscular dystrophy (BMD; OMIM #300376), owing to the formation of internally truncated, yet partially functional, dystrophin molecules. Hence, **Chapter 6** concerns investigations on the resetting of defective *DMD* reading frames in human myogenic cells by using HC-AdV delivery of the aforementioned optimized and multiplexing prime editing components. Finally, research described in **Chapter 6** establishes that combining straightforward HC-AdV transductions with seamless prime editing allows for stacking chromosomal edits in target cell populations through successive delivery rounds.

Similar to prime editors, base editors permit installing specific base pair changes in the genome requiring in the process neither DSB formation nor donor DNA delivery. The main base editing platforms consist of cytidine base editors (CBEs) and adenine base editors (ABEs) with CBEs and ABEs yielding C•G-to-T•A (C→T) and A•T-to-G•C (A→G) transitions, respectively. Importantly, although prime editors and base editors both constitute powerful high-potential tools for genetic therapies, their performance and precision at alternate chromatin states governing cell differentiation and identity, remain ill-defined. To address this knowledge gap, in **Chapter 7**, complementary loss-of-function and gain-of-function cellular systems are implemented to provide in-depth information concerning the efficiency and fidelity attained by using prime editors and base editors at euchromatin versus heterochromatin. The resulting findings inform and help guiding the development, selection and application of these powerful tools in specific cell types and contexts.

Taken together, the research presented in this thesis expands the current knowledge and toolbox underlying genome editing procedures through a comprehensive investigation of fast-developing genome editing systems and strategies in different cellular contexts. In particular, it reveals the feasibility and utility of using regular and high-specificity nicking RGNs for achieving efficient and accurate genetic modification of human cells involving targeted gene knockouts and HDR-mediated gene knock-ins. Moreover, it also establishes the suitability of the HC-AdV platform for the versatile investigation of advanced prime editing systems independently of their size and component numbers, which should facilitate the screening and application of the said systems in basic science and biotechnological settings. Finally, this thesis establishes causal relationships between specific chromatin states and the activities and fidelities attained by base editing and prime editing complexes in human cells, which has consequences for their further development and optimal deployment.

Chapter 1

Review

Adenoviral vectors meet gene editing: a rising partnership for the genomic engineering of human stem cells and their progeny

Francesca Tasca[1,†], Qian Wang[1,†] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands
[†]These authors contributed equally

Cells, 2020, Vol.9, No.4, ***953***
doi: 10.3390/cells9040953

ABSTRACT

Gene editing permits changing specific DNA sequences within the vast genomes of human cells. Stem cells are particularly attractive targets for gene editing interventions as their self-renewal and differentiation capabilities consent studying cellular differentiation processes, screening small-molecule drugs, modeling human disorders, and testing regenerative medicines. To integrate gene editing and stem cell technologies, there is a critical need for achieving efficient delivery of the necessary molecular tools in the form of programmable DNA-targeting enzymes and/or exogenous nucleic acid templates. Moreover, the impact that the delivery agents themselves have on the performance and precision of gene editing procedures is yet another critical parameter to consider. Viral vectors consisting of recombinant replication-defective viruses are under intense investigation for bringing about efficient gene-editing tool delivery and precise gene-editing in human cells. In this review, we focus on the growing role that adenoviral vectors are playing in the targeted genetic manipulation of human stem cells, progenitor cells, and their differentiated progenies in the context of in vitro and ex vivo protocols. As preamble, we provide an overview on the main gene editing principles and adenoviral vector platforms and end by discussing the possibilities ahead resulting from leveraging adenoviral vector, gene editing, and stem cell technologies.

1. INTRODUCTION

1.1. The Main Gene Editing Principles Based on Programmable Nucleases and Their Key Pros and Cons

Commonly, gene editing is triggered after programmable nucleolytic enzymes bind to predefined chromosomal sequences and locally generate double-stranded or single-stranded DNA breaks (DSBs or SSBs, respectively). The ensuing mending of these chromosomal breaks by cellular DNA repair mechanisms leads to the installation of targeted genomic changes whose extent can span from single to thousands of base pairs (bps).

Gene editing endeavors can disable a coding sequence (knockout) or remove specific genomic tracts. Moreover, they can equally restore a coding sequence or insert into specific genomic locations new genetic information (knock-in) present in exogenous (donor) DNA molecules. Typically, DNA editing strategies that knock-out or restore endogenous coding sequences involve the transfer of programmable nucleases that generate frameshifting insertions and deletions (indels) after the repair of targeted DSBs by non-homologous end joining (NHEJ) pathways. These include, classic NHEJ (cNHEJ) and alternative NHEJ (alt-NHEJ) pathways such as microhomology-mediated end-joining (MMEJ) and single-strand annealing (SSA) (1). The cNHEJ is the most active and fast-acting of the DNA repair pathways in mammalian cells often resulting in no or limited end-processing by exonucleases prior to ligation of chromosomal ends (1). Importantly, chromosomal ligation products containing indels can be generated (1), especially in the presence of a programmable nuclease that re-cleaves precisely ligated products until an indel disrupts its target site and becomes “fixed” in the cell population. It is also noteworthy mentioning that; (i) the target site sequences, (ii) the class of programmable nuclease employed, and (iii) the type of repair mechanism engaged in DSB repair, all contribute to different indel profiles which vary considerably in length and nucleotide composition (1,2). Yet, depending to some extent on microhomologies, the targeting of specific sequences by a programmable nuclease can yield specific indels in a high frequency of modified alleles (3-7).

Indels resulting from NHEJ-mediated repair of targeted DSBs can be exploited for disrupting non-coding elements (e.g., splicing motifs to induce exon-skipping) or reframing coding sequences that rescue endogenous gene expression via bypassing preexisting nonsense mutations (i.e., premature stop codons) (8,9). Alternatively, indels can be exploited for disrupting coding sequences that knockout endogenous gene expression via installing stop codons that induce nonsense-mediated mRNA decay (NMD) (9-11). However, it is important to mention recent research demonstrating the existence of an evolutionary conserved NMD-dependent mechanism in which the presence of a nonsense mutation in a gene can activate transcription of related genes whose products functionally complement the mutant gene (12,13). Another cautionary note concerns other recent findings in which DSB-derived indels in coding sequences can generate transcripts yielding various types of aberrant gene products (14). Therefore, these recently characterized processes, involving either genetic compensation responses triggered by indel-derived nonsense mutations or indels as such, have the potential of hindering the creation of robust gene knockout phenotypes and predictable gene editing outcomes. For a more

thorough and predictable removal of pre-existing genetic information, so-called multiplexing gene editing approaches can be deployed instead. In this case, two different programmable nucleases work in concert to generate a pair of intrachromosomal DSBs that lead to the excision of the intervening DNA sequence after end-to-end NHEJ ligation of the chromosomal termini (9,15-17). Alternatively, two programmable nucleases designed for generating a pair of inter-chromosomal DSBs can direct the assembly of specific translocations to, for instance, confirm or study the involvement of these translocations in cellular transformation events and, ultimately, cancer emergence (9,18).

Normally, knocking-in gene editing strategies encompass the delivery of programmable nucleases together with exogenous donor DNA that is inserted at the site-specific DSB via either homology-independent pathways (e.g., NHEJ) (19) or homology-directed DNA repair (HDR) (9-11). Generally, HDR-mediated knock-ins are more precise than those resulting from homology-independent processes in that they lack extraneous footprints at the border between endogenous and exogenous DNA. Indeed, instead of direct exogenous-to-endogenous DNA ligations via NHEJ or MMEJ, whose junction products often contain differently sized indels or specific footprints, DSB repair through HDR is a higher fidelity process (1,20). This process involves genetic exchange between donor and target sequences and includes extensive exonucleolytic processing of chromosomal breaks, single-strand invasions, and DNA synthesis over DSB-repairing donor templates (20). Ultimately, these molecular interactions result in accurate "copy-pasting" of the foreign genetic information into a specific locus (9-11). Yet, HDR-mediated gene editing is normally less frequent than gene editing based on DNA repair mechanisms that are independent of large tracts of homology between target and donor DNA templates (e.g., cNHEJ and MMEJ). In fact, as aforementioned, cNHEJ is the main DSB repair mechanism in mammalian cells (1,20). Further contributing to the differences in knocking-in frequencies obtained through gene editing involving cNHEJ versus HDR is the fact that the former pathway is active throughout the cell cycle; whereas the latter is only operative during the S and late G2 phases, when normally sister chromatids are available as sources of endogenous DNA-repairing templates (1,20). For this reason, gene editing involving the recruitment of the HDR pathway is unsuitable in non-cycling cells, such as, quiescent human hematopoietic stem cells (hHSCs) and terminally differentiated cells. Another consideration concerns the steep decline in HDR-mediated gene editing frequencies as the length of the exogenous DNA increases and the extent of continuous homology between target and donor DNA decreases (21). Therefore, the choice of the DSB repair pathway to exploit, and hence the designing of the DSB-repairing substrates to use, is contingent upon the specific application(s). For instance, knocking-in large genetic payloads into introns of safe harbor loci (e.g., *AAVS1* and *CCR5*) for achieving stable and homogeneous transgene expression in cell populations may be best pursued via selecting HDR-independent gene editing strategies; whereas knocking-in donor DNA into coding sequences for modeling or repairing genetic defects in stem or progenitor cells is best accomplished through precise HDR-dependent gene editing.

1.2. The Main Programmable Nuclease Platforms and Their Key Pros and Cons

Under regular conditions, HDR-mediated gene knock-ins are very rare events in human cells, with typical frequencies varying between 10^{-6} and 10^{-7} (22–24). The finding that site-specific DSBs made by homing endonucleases at chromosomally embedded recombinant sequences could stimulate HDR by several orders of magnitude, was a powerful stimulus for the development of programmable nucleases (25–27).

The crucial feature of programmable nucleases is their capability of binding to and cleaving at predefined DNA sequences, including those located within large genomes (9,10,11,28). Nowadays the main classes of programmable nucleases are, in chronological order of appearance, zinc-finger nucleases (ZFNs) (29), transcription activator-like effector (TALE) nucleases (TALENs) (30-34), and RNA-guided nucleases (RGNs) (35-38). Naturally, the development of programmable nuclease technologies was invariably grounded on fundamental insights obtained from a broad range of biological systems, spanning from vertebrate cells and phytopathogenic bacteria, in the case of ZFNs (39) and TALENs (40,41), respectively, to bacteria and archaea, in the case of RGNs (42,43).

ZFNs and TALENs are modular proteins that present an overall similar architecture (**Figure 1A** and **B**). In particular, they consist of a customizable DNA-binding domain fused through a flexible linker to a non-specific nuclease domain, typically that of the type IIS FokI restriction enzyme whose catalytic

activity is dependent on dimerization (44). Resulting from their comparable generic architectures, ZFNs and TALENs act in a similar fashion in that members of ZFN and TALEN pairs bind in close proximity to each other on opposite DNA strands of a bipartite target sequence leading to site-specific DSBs at the spacer region after local dimerization of the FokI nuclease domains (**Figure 1A** and **B**). The DNA-binding domains of ZFNs and TALENs consist of arrays of engineered zinc-finger motifs and TALE repeats, respectively, with each zinc-finger motif usually binding to nucleotide triplets and each TALE repeat binding to single nucleotides within their respective double-stranded target sites (**Figure 1A** and **B**). Cys_2-His_2 zinc-fingers are found in metazoans where they serve as motifs in RNA and DNA binding proteins whose wide roles include transcriptional and epigenetic regulation of target genes (45,46). Native TALE proteins are found in certain phytopathogenic bacteria (e.g., *Xanthomonas sp.*) where they serve as virulence factors once injected into host plant cells via type III secretory apparatuses (47). The binding of zinc-finger motifs to specific triplets can be substantially affected by flanking nucleotides (48). This sequence context dependency contributes to making highly specific ZFNs a laborious task requiring complex protein engineering methodologies that may include several rounds of optimization and/or screening and selection of ZFN candidates from large zinc-finger libraries (48). In contrast, the binding of TALE repeats to their cognate nucleotides does not seem to be substantially influenced by neighboring sequences (49). This limited sequence context dependency aids the assembly of functional and highly specific TALENs whose designing flexibility and genomic space coverage is superior to that of ZFNs (49). DNA binding of TALEs are, however, significantly hindered by cytosine methylation (50,51) and Krüppel-associated box-induced heterochromatin (52). Importantly, the former epigenetic modification can be elegantly surpassed by incorporating non-canonical TALE repeats within TALE arrays (51).

Native RGNs are found in many bacteria and archaea where they form adaptive immune systems against invading agents, e.g., bacteriophages and foreign plasmids (53). Engineered RGNs, such as those based on the prototypic clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated 9 (Cas9) system, from *Streptococcus pyogenes* (35-38), operate differently from ZFNs and TALENs in that target DNA cleavage does not depend exclusively on protein-DNA binding but also on RNA-DNA hybridization. In particular, RGNs, consisting of a sequence-specific single guide RNA (gRNA) coupled to an invariant nuclease, first recognize so-called protospacer adjacent motifs (PAMs) on the DNA via PAM-interacting domains in the nuclease component (10,54). In the case of the *S. pyogenes* Cas9 the PAM reads NGG. Typically, in instances in which the 19-21 deoxyribonucleotides "upstream" from the PAM are complementary to the 5' end the gRNA, DSB formation ensues through the concerted action of the HNH and RuvC-like nuclease domains of Cas9 (**Figure 1C**). The events leading to DSB formation upon initial Cas9-PAM interrogation include, PAM-proximal DNA unwinding, R-loop formation and expansion via increasing gRNA:DNA annealing which subsequently triggers HNH translocation and pairing with the RuvC-like domain. Ultimately, HNH-RuvC pairing catalyzes phosphodiester bond cleavage of both DNA chains, predominantly three base-pairs upstream from the PAM (**Figure 1C**) (10,43,55).

Crucially, RGNs can cut DNA at unintended genomic sequences (off-target sites) especially if mismatches between gRNA and DNA sequences locate at PAM-distal positions (56-60). Furthermore, albeit to a lesser extent than NGG, *S. pyogenes* Cas9 can also effectively engage non-canonical PAMs (e.g., NAG), which further contributes to off-target activities (57,60,61). Therefore, similarly to their programmable nuclease predecessors, the application of RGNs warrants careful assessment of potential off-target sites, especially if directed toward clinical testing. Indeed, judiciously chosen gRNAs can, per se, greatly reduce off-target activities in vitro and in vivo (62,63). As TALENs, targeted DNA cleavage by RGNs is also hindered to some extent by epigenetic mechanisms underpinning specific heterochromatic states (52,64-66). However, in contrast to TALENs, RGNs do not seem affected by DNA methylation (57).

The fact that readdressing RGNs to new target sites simply comprises modifying the 5' end of the gRNA component, and hence does not require de novo protein engineering as ZFNs and TALENs do, confers these CRISPR-based nucleases with unsurpassed versatility and ease-of-use. Such features have fueled the primacy of RGNs amongst current programmable nuclease platforms. In fact, since the initial adaptation of natural CRISPR-Cas9 systems into genome engineering tools (35-38), RGN technologies are diversifying, being combined and adapted, at increasing rates (67). For instance, structure-guided

rational design and directed evolution approaches are producing new Cas9 variants whose features include; recognition of alternative PAMs that broaden the range of targetable genomic sites and improved target site specificities (67). In parallel with these developments, phylogenetic analyses and mining of metagenomic datasets are unearthing components that make-up the highly diverse universe of CRISPR systems which, in addition to DNA, also target and degrade invading RNA (53). Many of these components end up being successfully converted into reagents for (epi)genome and transcriptome modification or modulation in mammalian cells (67-69).

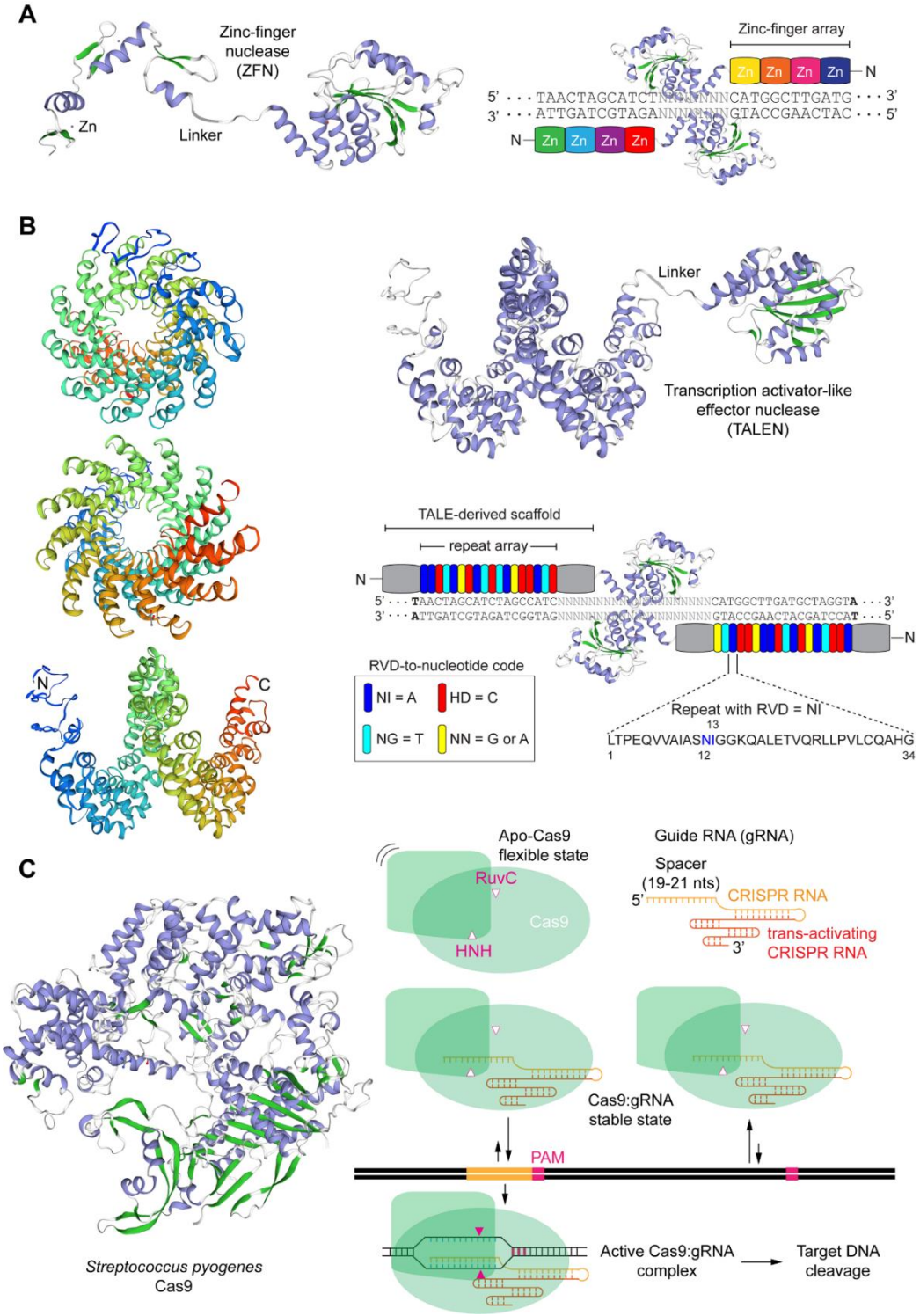


Figure 1. Schematics of the main programmable nuclease platforms. (**A**) Zinc-finger nucleases (ZFNs). ZFNs are chimeric modular DNA-binding proteins consisting of the FokI nuclease domain fused through a flexible linker to an array of 3–6 artificial Cys_2-His_2 zinc-finger motifs. Each zinc-finger motif acquires its structure through tetrahedral coordination of 2 cysteines in β-

sheets and 2 histidines in α-helixes by zinc ions. ZFN monomers of a working ZFN pair bind on opposite DNA strands in a tail-to-tail configuration leading to local FokI nuclease domain dimerization and ensuing site-specific double-stranded DNA breaks (DSB) formation within the spacer sequence. (**B**) Transcription activator-like effector (TALE) nucleases (TALENs). TALENs are chimeric modular DNA-binding proteins comprising the FokI nuclease domain fused through a flexible linker to a series of typically 17.5 repeats derived from TALE proteins. TALE proteins contain a translocation and transcriptional activation domain separated by a central array of typically 33-35 isomorphic repeats. The repeats harbor at amino acid positions 12 and 13 highly polymorphic residues named repeat variable di-residues (RVDs) that bind to specific nucleotides. The structure of 17.5 TALE repeats from an engineered TALEN monomer are depicted in frontal and lateral views. TALEN monomers of a working TALEN pair bind on opposite DNA strands in a tail-to-tail configuration resulting in local FokI nuclease domain dimerization and ensuing site-specific DSB formation within the spacer sequence. (**C**) RNA-guided CRISPR-Cas9 nucleases. Engineered CRISPR-Cas9 nucleases are sequence-specific ribonucleoprotein complexes consisting of a Cas9 protein with two nucleases domains (i.e., HNH and RuvC-like) bound to a single guide RNA (gRNA) formed by a sequence customizable CRISPR RNA (crRNA) fused to a constant trans-activating CRISPR RNA (tracrRNA) scaffold moiety to which the *S. pyogenes* Cas9 enzyme binds to. Target sequences of Cas9:gRNA complexes consist of the protospacer-adjacent motif (PAM) NGG placed next to an usually 20 nucleotide-long sequence complementary to the 5′-terminal end of the crRNA (spacer). The tertiary protein structures shown, each of which derived from the primary amino acid sequences of specific ZFN, TALE and Cas9 reagents, were homology-modeled through the SWISS-MODEL server. β-sheets and α-helixes are colored in green and violet, respectively.

1.3. A Brief Overview on the Biology of Adenoviruses and Their Recombinant Types

Adenoviruses are a diverse group of viruses from the Adenoviridae family that have been evolving in a wide range of vertebrates, including humans, where they cause mild ailments, e.g., in the respiratory and gastrointestinal tracts (70-72). Human adenoviruses belong to the *Mastadenovirus* genus with over 55 different serotypes identified so far. The various serotypes are grouped in species A through G based on phylogenetic, genome structure and hemagglutination criteria. Structurally, adenoviral particles (virions) consist of a non-enveloped icosahedral protein capsid displaying protruding fibers (70-72) (**Figure 2**). A linear double-stranded DNA genome with terminal proteins bound to their 5′ ends is packaged inside each virion capsid consisting of 240 trimers of the hexon protein, 12 pentamers of the penton base polypeptide and 12 trimeric fiber proteins that protrude from each of the 12 capsid vertices (**Figure 2**). Each homo-trimeric fiber consists of a basal tail domain that docks within the penton base axis, a slender shaft region and an apical globular knob domain responsible for the initial attachment of the virion to host cell receptors (**Figure 2**). In addition to the major capsomers hexon, penton base, and fiber, the adenoviral capsid also contains other so-called minor proteins some of which are thought to be important for cementing the virion structure (72,73). Adenovirus serotypes present broad cellular tropisms owing to their usage of a wide range of cell surface receptors. Identified primary attachment receptors include, the coxsackie and adenovirus receptor (CAR) used by the prototypic serotypes 2 and 5 from species C (74,75) and CD46 and desmoglein-2 engaged by species B serotypes (76,77). Certain serotypes engage instead glycans and polysialic acids as primary attachment moieties (78,79). The natural diversity of adenoviruses and their corresponding wide range of host-cell receptors is permitting; (i) constructing new vectors based on rare serotypes that can escape pre-exiting immunity to adenoviruses prevalent in the human population, for anti-cancer and vaccination purposes (80); and (ii) changing the tropism of established vectors based on species C adenovirus serotype 5 into those of other serotypes so that cells with therapeutic relevance lacking CAR can be efficiently transduced (81). For instance, genetic retargeting of vector particles through the exchange of the apical regions of the adenovirus serotype 5 fiber (**Figure 2**) for those of species B adenovirus serotype 35 or 50 permits efficient transduction of $CAR^{low}/CD46^{high}$ hHSCs (82,83), human mesenchymal stromal cells (hMSCs) (84,85) and human muscle progenitor cells (86).

The processes through which adenoviruses introduce their genomes into host-cell nuclei have been most extensively studied in the case of serotype 5 (87). Briefly, after the initial attachment to the host cell, endocytosis via clathrin-coated vesicles is triggered by interactions between RGD motifs in penton bases and cellular integrins (e.g., $\alpha_v\beta_5$). Subsequently, incoming fiberless virions escape lysosomal degradation via the lowering of the pH in endosomes that permits remodeled capsid components to lyse the vesicle membranes. Once in the cytosol, the remodeled nucleocapsids bind to motor proteins dynein/dynactin that transports them along the microtubule network until they dock at the nuclear pore complex and release the packaged DNA into the nucleoplasm (87).

The most thoroughly used adenoviral vectors (AdVs) are deleted in the transcriptional units *E1A* and *E1B* that make-up the early region 1 (*E1*) (**Figure 3**). The production of these first-generation, *E1*-deleted, AdVs takes place in packaging cell lines (e.g., HEK293 and PER.C6) that express, and hence complement, *in trans* the *E1* gene products (88,89). The deletion of *E1*, firstly, blunts the activation of

the regular adenoviral gene expression program preventing the replication of vector particles in transduced cells and, secondly, creates room for the packaging of approximately 5.0 kb of exogenous DNA in adenoviral capsids. Since the *E3* region is dispensable for replication in cell culture systems, some vector designs combine deletions in *E1* with deletions in *E3* that permit the packaging of up to 8 kb of exogenous DNA (71). As it came to be known, the *E1* deletion does not fully prevent residual expression from some of the transcriptional units that remain in vector genomes (71). The resulting leaky synthesis of viral gene products leads to vector dose-dependent cytotoxicity in vitro and short-lived transgene expression in vivo (2-3 weeks) due to the clearance of transduced cells by the immune system (90). For this reason, *E1*-deleted AdVs, in particular those based on serotypes with low seroprevalence in the human population, are being applied in clinical trials not for gene therapies requiring prolonged transgene expression but as vaccination agents instead, e.g., against hemorrhagic fever and AIDS caused by Ebola and HIV-1 infections, respectively (91,92).

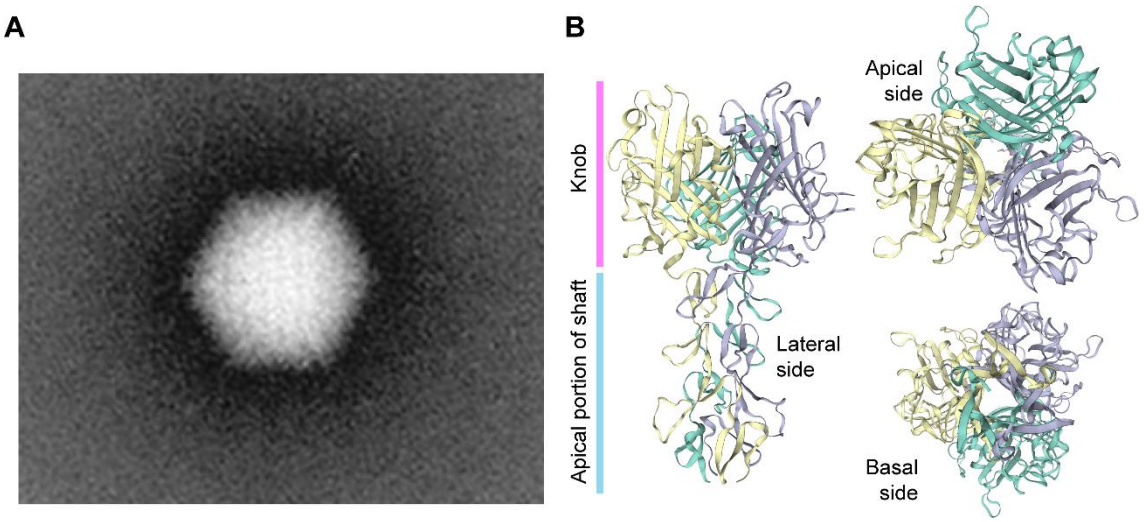


Figure 2. Adenovirus particle and the structure of its cell receptor-interacting fibers. (**A**) Transmission electron microscopy image of an adenovirus particle (virion). The icosahedral shape of the non-enveloped virion capsid can be discerned (~90 nm). A few of the twelve slender protruding fibers with their apical globular knob domains responsible for the initial interaction with the host-cell coxsackie and adenovirus receptor (CAR), can equally be discerned. (**B**) Three-dimensional model of the apical regions of the adenovirus serotype 5 fiber. The fiber is a homotrimer of the polypeptide encoded by the *L5* open reading frame and consists of the tail (not shown), the rod-like shaft and the globular knob domains. The tail anchors the fiber to the adenovirus capsid via non-covalent binding to the penton base proteins; the shaft projects the knob away from the capsid facilitating its interaction with CAR on the surface of host cells. The quaternary protein structure was homology-modeled using the SWISS-MODEL server and is depicted in different angles.

Second-generation AdVs combine deletions in *E1* or *E1* and *E3* with deletions in other early regions, i.e., *E4* or *E2* (**Figure 3**). Therefore, these vectors are generated in specialized packaging cell lines that complement in trans the respective missing gene products (71). Although second-generation AdVs are more crippled than first-generation AdVs, at high vector doses, leaky synthesis of viral gene products can still be detected which also correlates with short-term transgene expression in vivo (71,93).

To abrogate altogether leaky viral gene expression in transduced cells and, at the same time, maximize the size of foreign DNA that can be incorporated in adenoviral capsids, high-capacity adenoviral vectors (HC-AdVs) were developed (71) (**Figure 3**). These third-generation AdVs (a.k.a. “gutless” or helper-dependent AdVs) lack all viral coding sequences retaining from the parental virus genome exclusively the short cis-acting inverted terminal repeats (ITRs) (103-bp each) and packaging elements needed for, respectively, vector DNA replication and encapsidation in producer cells (**Figure 3**). The need for complementing *in trans* the full set of adenoviral gene products, makes the production of HC-AdVs more complex than that of their earlier generation counterparts. In particular, HC-AdV particles are assembled in *E1*-complementing cell lines that express a site-specific recombinase (e.g., Cre or FLP) (71,94,95). These producer cell lines are transduced with an *E1*-deleted helper AdV that expresses in trans the viral gene products necessary for the replication and packaging of HC-AdV genomes into adenoviral capsids. Crucially, the packaging signals of the helper genomes are flanked by recognition sequences for the site-specific recombinase so that the vast majority of assembled AdV capsids contain HC-AdV DNA in detriment of helper DNA owing to the selective recombinase-mediated removal of the packaging

elements from the latter templates. Normally, besides the adenoviral cis-acting elements and the foreign DNA of interest, HC-AdV genomes also contain a so-called “stuffer” DNA segment to increase the HC-AdV DNA length to at least ~28 kb and, in doing so, guarantee vector genome stability during replication in producer cells (94).

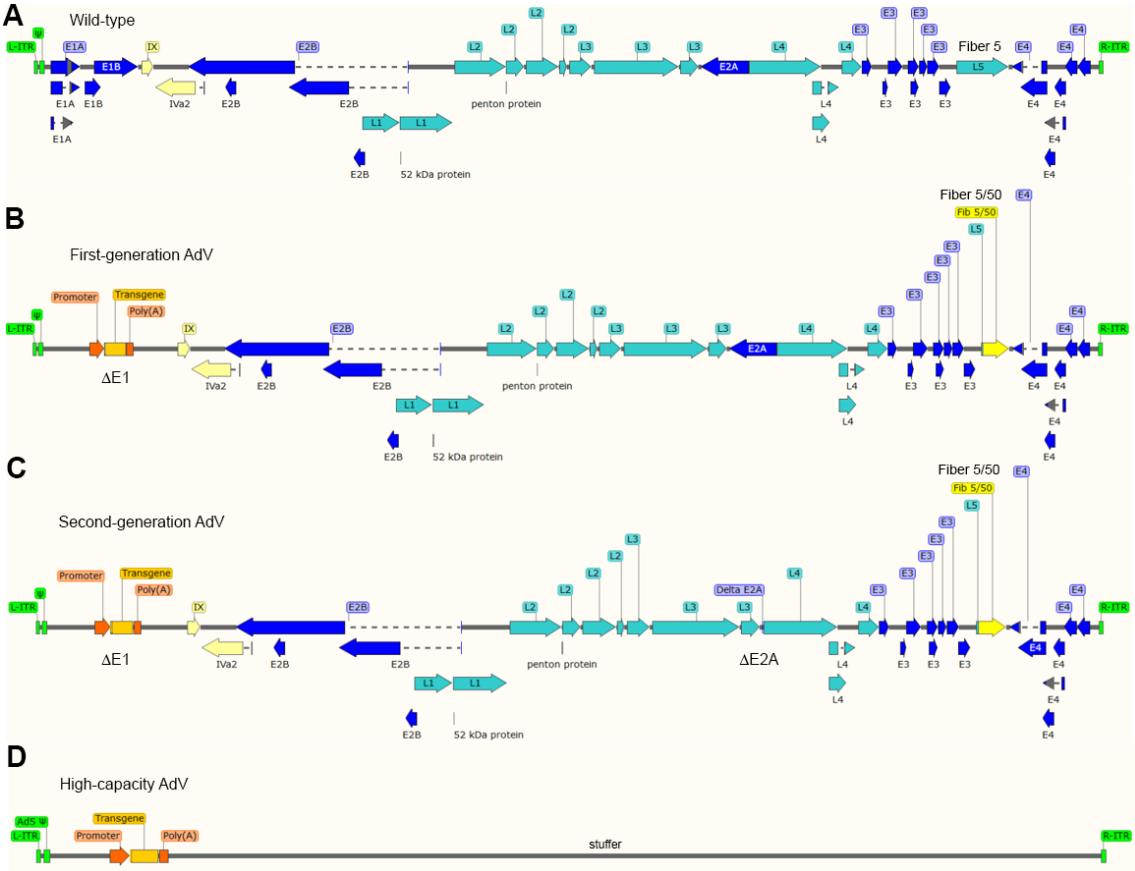


Figure 3. Schematics of wild-type and recombinant adenoviruses. (**A**) Genome structure of the prototypic human adenovirus serotype 5 drawn in relation to the genome structures of (**B**) first-generation (*E1*-deleted), (**C**) second-generation (*E1*- and *E2A*-deleted), and (**D**) third-generation or high-capacity (fully viral gene deleted) adenoviral vectors. The vectors contain a typical expression unit (transgene) consisting of a coding sequence of interest under the transcriptional control of a heterologous promoter and a polyadenylation signal. The first- and second-generation vector genomes encode chimeric fibers consisting of the basal shaft sequence of the human adenovirus serotype 5 linked to the apical shaft and knob domains from the CD46-interacting human adenovirus serotype 50 (yellow arrows). The non-coding *cis*-acting elements involved in vector genome replication and encapsidation are the inverted terminal repeats (ITRs) and packaging signal (Ψ), respectively. The latter signal and the “left” and “right” ITRs (L-ITR and R-ITR, respectively) are depicted in green. Regulatory functions necessary for activating the viral gene expression program are encoded by the early (E) regions *E1A*, *E1B*, *E2A*, *E3* and *E4* (dark blue arrows). The structural proteins required for assembling mature virions are encoded by the late (L) regions *L1* through *L5* (light blue arrows). The *L5* open reading frame (ORF) yields the cell surface receptor-interacting fibers. The full activation of the late viral gene expression program takes place after the onset of viral DNA replication. The ORFs coding for the intermediate proteins IX and IVa2 are also shown (light yellow arrows). Other adenoviral ORFs, e.g., small non-coding RNAs VAI and VAII are not depicted. The SnapGene software (version 5.0.7) was used for generating the different diagrams on the basis of the human adenovirus serotype 5 source sequence retrieved from GenBank accession number: AY601635.1.

2. Adenoviral Vector-Based Gene Editing in Human Adult Stem Cells and Their Progeny

2.1. Targeted Gene Disruption

Various viral vector systems initially developed for transgene expression and gene therapy undertakings, have also started to be investigated and coopted as gene editing agents (for a review on their features and main pros and cons, see, ref. 9). In fact, all three classes of replication-defective AdV systems (**Figure 3**) are included in these gene-editing research efforts, that are covered next.

E1-deleted AdVs based on serotype 5 displaying apical fiber motifs from CD46-interacting serotype 35 (AdV5/35) have been tested for conferring resistance to HIV-1 infection. In particular, AdV5/35 vectors encoding *CCR5*-specific ZFNs were applied for NHEJ-mediated generation of human $CD4^+$ T cells with reduced amounts of the transmembrane HIV-1 co-receptor protein CCR5 (96). The ex vivo cell

transduction protocol resulted in 40–60% disruption of *CCR5* alleles in these cells. Importantly, transplantation experiments in immunodeficient NOD/Shi-scid/$\gamma c^{-/-}$ (NOG) mice led to a 3-fold enrichment of $CD4^+$ T cells with *CCR5* knockout alleles in animals infected with HIV-1, suggesting selection for gene-modified cells. Notably, next generation sequencing (NGS) analysis of transduced $CD4^+$ T cells revealed a substantial ZFN-induced off-target activity (i.e., 5.39% indels) at the neighboring and highly sequence identical *CCR2* locus (96). Building on this principle but aiming at a longer protective effect against HIV-1 infection, another study focused on targeting adult hematopoietic stem/progenitor cells (HSPCs). In this work, AdV5/35-mediated delivery of *CCR5*-specific ZFNs into HSPC-enriched $CD34^+$ cells led to target allele knockout frequencies above 25%. However, these knockout levels were only obtained in the presence of protein kinase C (PKC) activators, an expedient used to presumably improve vector transduction and/or ZFN expression (97). Moreover, low, yet detectable, off-target activity at *CCR2* and at three other non-coding sequences located elsewhere in the genome were observed via NGS analysis. Subsequent cell transplantation assays in immunodeficient NOD/SCID/$\gamma c^{-/-}$ (NSG) mice showed a vector dose-dependent reduction in the levels of human cell engraftment as measured by $CD45^+$ cell counts in animals infused with HSPCs treated with PKC activators and *CCR5*-targeting ZFNs (97). To avoid the toxicity caused by PKC activators, Maier and co-workers tested instead anti-CD3/CD28 stimulation as an adjuvant for improving transduction of T lymphocytes by an AdV5/35 vector encoding *CCR5*-specific ZFNs (98). When compared to the experimental group exposed to PKC activation, this method enhanced the frequencies of target gene knockout by almost 3-fold (up to 32%). Importantly, ZFN-associated toxicity was not detectable with levels of off-target *CCR2* disruption in transduced T lymphocytes remaining below 4%, as estimated through genotyping assays based on mismatch-sensing nucleases and DNA fluorescence densitometry (98).

The generation of AdVs encoding ZFNs is challenging due to cytotoxicity caused by transgene overexpression in producer cells. To overcome this limitation, Saydaminova and colleagues exploited miRNA-dependent downregulation of transgene expression in 293-Cre packaging cells. This strategy permitted generating tropism-modified HC-AdVs encoding *CCR5*-specific ZFNs at high yields and without vector genome rearrangements. Importantly, miRNA profiling guaranteed that the endogenous miRNA suppressing ZFN synthesis in producer cells was not expressed in hHSC-enriched $CD34^+$ target cells (99). Transduction of the erythroleukemia cell line MO7e and primary $CD34^+$ cells with the resulting HC-AdV coding for the *CCR5*-specific ZFNs led to 43.6% and 13% indel formation, respectively, at *CCR5* as determined by mismatch-sensing nuclease assays. Cell transplantation experiments in immunodeficient NOG mice revealed, however, that human $CD34^+$ cells transduced with the ZFN-encoding HC-AdV engrafted in the bone marrow at 3-fold lower levels than their non-transduced counterparts (i.e., 2.12% versus ~6%, respectively) (99).

A *CCR5*-specific ZFN pair delivered ex vivo into autologous $CD4^+$ T cells of AIDS patients by an *E1*-deleted AdV5/35 vector formed the basis for the first clinical testing of a programmable nuclease (100). The infusion of 10 billion cells, of which 11–28% were *CCR5*-disrupted, was shown to be safe. Moreover, edited cells persisted after transplantation with a mean half-life of 48 weeks and, tantalizingly, upon an interruption of anti-retroviral therapy, the rates with which *CCR5*-disrupted cells declined were significantly slower than those of unmodified cells (100). Outstanding questions following from this landmark study are the feasibility in achieving sufficient numbers of cells with bi-allelic *CCR5* knockout without inducing cytotoxicity and with minimal ZFN-induced off-target effects. Finally, the combination of genetically retargeted AdV5/35 vectors and ZFN technologies has also been used for knocking out endogenous T-cell receptor genes and the primary HIV-1 receptor gene *CXCR4* in T cells (101,102).

In addition to ZFNs, the AdV platform is equally suitable for the delivery of TALENs into human somatic cells, e.g., muscle progenitor cells and hMSCs. In fact, Holkers and co-workers demonstrated that, in contrast to HIV-1-based lentiviral vectors, transgenes encoding TALENs can be transferred intact into human cells by AdVs (103). Indeed, lentiviral vectors encoding TALENs suffer substantial genetic rearrangements in the form of deletions of various sizes that occur within the direct repeats corresponding to the TALE DNA-binding domains (**Figure 1B**). These deletions are likely caused by frequent reverse transcriptase template switching events taking place within the TALE repetitive tracts. Thus, the transfer of transgenes coding for TALE-based proteins through standard and integration-

defective lentiviral vectors (IDLVs) requires substantial coding sequence optimization for minimizing sequence identity among repeats (104,105).

It is also noteworthy to mention that, although IDLVs permit transient expression of ZFNs and sequence optimized TALENs in human cells, the yields necessary for robust targeted DSB formation might not be reached due to epigenetic silencing mechanisms directed at IDLV genomes involving histone deacetylases (106,107). In contrast, functional assays revealed that AdVs expressing TALENs allow for robust targeted DSB formation in several human cell types, e.g., muscle progenitor cells and hMSCs (103). Second-generation AdVs deleted in *E1* and *E2A* and displaying apical motifs from CD46-interacting serotype 50 (AdV5/50) were used in these proof-of-concept experiments validating the AdV platform for the delivery of functional TALENs into human cells (103). Follow-up experiments using first-generation and second-generation fiber-modified AdVs encoding TALENs and *S. pyogenes* Cas9 addressed to sequences flanking the major *DMD* mutational hotspot triggered large deletions comprising multiple exons (>500 kb) in patient-derived muscle progenitor cells (17). These maneuvers designed for repairing *DMD* alleles causing Duchenne muscular dystrophy (DMD), led to the synthesis of in-frame mRNA transcripts encoding a truncated yet potentially functional Becker-like dystrophin protein (17).

Currently, the integration of AdV and programmable nuclease technologies for gene editing in somatic cells is dominated by the delivery and testing of RGNs. The first viral vector-mediated delivery of RGN components into mammalian cells consisted of using fiber-modified *E1*- and *E2A*- deleted AdVs expressing Cas9 or gRNAs directed to either a chromosomally integrated *EGFP* reporter or to the *AAVS1* safe harbor locus located in the human chromosome 19 at position 19q13.3-qter. In co-transduction experiments, robust targeted DSB formation was achieved at *AAVS1* in several cell types including human muscle progenitor cells and hMSCs (108). In another study, co-transduction of human lung microvascular endothelial cells with an *E1*-deleted AdV and a lentiviral vector encoding Cas9 and a *TIE2*-specific gRNA, respectively, induced up to 90% of target gene disruption. Direct phenotypic analysis of *TIE2*-edited cell populations showed a persistent increase in endothelial cell permeability when compared to control cells (109).

In addition to NHEJ-mediated target gene disruption for basic biology studies, AdV-mediated RGN delivery is also being explored for modifying genes underlying human disorders. In this regard, to facilitate the delivery of RGN components, Maggio and colleagues co-packaged Cas9 and gRNA expression units within single particles of fiber-modified *E1*- and *E2A*-deleted AdVs (17). In these experiments, testing “all-in-one” AdV-mediated transfer of RGN components, *DMD* exons 51 and 53 were separately targeted for resetting the *DMD* reading frame in muscle progenitor cells derived from DMD patients (17). In a follow-up study, fiber-modified *E1*- and *E2A*-deleted AdVs encoding Cas9 and gRNA pairs targeting *DMD* introns 52 and 53 or introns 43 and 54 were assembled for triggering single or multiple exon deletions, respectively (110). The latter dual RGN-encoding vector permitted removal of the aforementioned major *DMD* mutational hotspot in up to 18% of target alleles in patient-derived muscle progenitor cells (110). More recently, fiber-modified HC-AdVs were applied for the delivery of optimized high-specificity dual RGNs equally targeting *DMD* introns 43 and 54. The transduction of muscle progenitor cells isolated from DMD patients with these CD46-targeting HC-AdV particles resulted in the removal of the major *DMD* mutational hotspot in up to 42% of target alleles resulting in the direct detection of Becker-like dystrophin synthesis in differentiated muscle cell populations (111).

A study by Li and coworkers documented over 30% indel formation at *CCR5* in $CD4^+$ T cells that had been pretreated with a PKC activator and subsequently selected for RGN expression after exposure to *E1*-deleted AdV5/35 particles encoding EGFP-tagged RGNs. Significantly, the authors obtained evidence for the acquisition of resistance of *CCR5*-edited $CD4^+$ T cells to two different HIV-1 strains in vitro (112).

Disruption of binding motifs for the *HBG* repressor protein BCL11A is a promising strategy to reactivate *HBG* expression and fetal γ-globin synthesis to complement the absence of functional adult β-globin in β-thalassemic and sickle cell disease (SCD) patients. In this regard, transduction of mobilized peripheral blood $CD34^+$ cells from healthy donors with fiber-modified HC-AdVs encoding *HBG*-specific RGNs led to around 20% of target motif disruption in these cells (113). Moreover, no indels were observed in the

top 10 candidate off-target sites, as assessed by mismatch-sensing nuclease assays and, importantly, the erythroid differentiation capability of the gene-edited hematopoietic progenitors was maintained (113). Cell transplantation assays in lethally irradiated immunodeficient mice revealed indel frequencies ranging from 19% to 25% at *HBG* alleles in human CD45$^+$ cells isolated from bone marrow at 10 weeks post-transplantation. Upon in vitro differentiation of these bone marrow-derived CD45$^+$ cells, the frequencies of γ-globin$^+$ cells were ~50% and ~27% in the transduced and non-transduced groups, respectively, as determined by flow cytometry (113). In addition, β-YAC/CD46 mice were also used in this study to overcome the known block on human erythrocytic lineage differentiation in NSG mice. β-YAC/CD46 mice contain a human DNA fragment encompassing the entire 82-kb human *β-globin* locus and express the human CD46 receptor which permits transducing mouse cells with HC-AdV particles displaying adenovirus serotype 35 fibers. Hence, this mouse model allows in vivo evaluation of *HBG* reactivation in mature circulating erythrocytes. Bone-marrow Lin$^-$ cells isolated from β-YAC/CD46 mice were transduced with the fiber-modified HC-AdVs encoding *HBG*-specific RGNs and were subsequently transplanted into lethally irradiated C57BL/6 recipient mice. At 10 weeks post-transplantation, there was a ~5-fold reduction of *HBB* mRNA and a ~30-fold increase in *HBG* mRNA levels in red blood cells when compared to controls. These results indicate that a switch in the balance of adult to fetal globin expression was achieved (113). In another study, Li and co-workers using fiber-modified HC-AdVs encoding RGNs targeting *BCL11A* gene enhancer or BCL11A protein binding sequences obtained over 20% indel formation at these motifs in CD34$^+$ cells (114). Interestingly, however, in vitro colony-forming unit (CFU) assays based on semi-solid methyl-cellulose medium showed a reduction in the number of multi-lineage progenitors derived from vector-transduced cells (114). In addition, cell transplantation assays in irradiation-conditioned NSG mice demonstrated that engraftment rates of CD45$^+$ cells in mice receiving grafts transduced with RGN-encoding HC-AdVs were 5- to 10-fold lower than those transplanted with non-transduced cells or cells transduced with a control vector encoding exclusively Cas9 (114). The low numbers of CFUs in vitro and engraftment rates in vivo indicated RGN-induced cytotoxic effects. In line with this data, Schiroli and colleagues found through single-cell transcriptomics analysis that DSBs induced by ZFNs and RGNs can activate a P53-dependent DNA damage response in HSPCs (115). To shorten the duration of RGN activity, bacteriophage anti-CRISPR (Acr) peptides AcrIIA2 and A4, were exploited to inhibit long-term Cas9 activity (114). Sequential transfer of *BCL11A* enhancer-specific RGNs and Acr peptides via tropism-modified HC-AdV transductions with an interval of 48 hours led to 37.9% indel formation in the human umbilical cord blood-derived erythroid progenitor cell line HUDEP-2 (114). Flow cytometry and qRT-PCR analyses showed a switch of *HBB* to *HBG* expression in the edited HUDEP-2 populations. After applying a similar sequential HC-AdV transduction protocol to CD34$^+$ cells followed by transplantation of vector-treated cells into irradiation-conditioned NSG mice, Li and coworkers observed comparable levels of CD45$^+$ cell engraftment in mice receiving non-transduced and vector-transduced cells. Indel frequencies at the *BCL11A* gene enhancer and BCL11A protein binding site ranged from 8.5% to 27% and from 10.5% to 21%, respectively, in CD45$^+$ cells isolated from bone marrow, as measured by mismatch-sensing nuclease assays. Finally, in vitro differentiation of isolated CD45$^+$ cells into erythroid cells, revealed a ~1.4-fold increase in the percentage of γ-globin$^+$ cells in the edited over the control groups (114).

2.2. Targeted Gene Integration

As aforesaid, HDR leads to precise genomic DNA editing in the presence of exogenous donor templates that can be designed for gene knock-ins, gene knockouts or gene correction. Therefore, AdVs are also being utilized for transferring programmable nucleases together with donor templates into human cells. In this context, Coluccio and colleagues combined AdV-mediated ZFN delivery with the transfer of donor HDR substrates in AdVs or IDLVs for testing homology-directed gene insertion in human keratinocytes (116). In this study, *AAVS1*-specific ZFNs were delivered by an *E1*-deleted AdV5/35 vector, whereas the donor, containing a reporter gene flanked by *AAVS1*-targeting homologous sequences, was transferred via either vesicular stomatitis virus glycoprotein G-pseudotyped IDLV or *E1*-deleted AdV5/50 particles. Transduction of HaCaT cells, a human keratinocyte cell line, with ZFN-encoding AdV particles together with IDLV or AdV donors led to chromosomal transgene integration frequencies of 20% and 1%, respectively (116). However, combining AdV5/35 and IDLV vectors for introducing into human primary keratinocytes *AAVS1*-specific ZFNs and donor templates, respectively, resulted in substantially lower frequencies of stable transgene insertion (i.e., 0.3%), presumably in part due to the observed inefficient transduction of these target cells by IDLV particles (116). In another study, investigating homology-directed gene targeting, Holkers and coworkers combined the transfer of HDR substrates in

AdV or IDLV particles with AdV-mediated delivery of TALENs instead (117). In particular, *AAVS1*-specific TALENs were delivered by an *E1*-deleted AdV5/50 vector, whereas the donor, containing a reporter gene flanked by *AAVS1*-targeting sequences, was transferred via either IDLV or *E1*- and *E2A*-deleted AdV5/50 particles. Transduction of human muscle progenitor cells with TALEN-encoding AdVs together with IDLV or AdV donors led to chromosomal transgene integration frequencies of 9.1% and 1.24%, respectively. These data together with that of Collucio and coworkers indicate that IDLV donors lead to higher frequencies of DSB-dependent gene knock-ins than those achieved by AdV donors. However, isolation of genetically modified muscle progenitor cells (n = 214 clones) followed by clonal analysis using junction PCR assays demonstrated that a large proportion of IDLV-modified cells contained random insertions (13.4%) or inaccurate *AAVS1* insertions (44.3%), of whom a substantial fraction corresponded to head-to-tail donor DNA concatemers (38.5%). In contrast, neither random insertions nor inaccurate *AAVS1* insertions were detected in the randomly isolated AdV-modified cells (117). Thus, although free-ended IDLV genomes lead to higher frequencies of genetically modified cells than protein-capped AdV genomes, the latter genomes result in more specific and accurate HDR-mediated donor DNA insertion (28,117). The relevance of the donor DNA structure to the specificity and accuracy of gene targeting was demonstrated by experiments in which the excision of HDR substrates from the context of protein-capped AdV genomes resulted in an increase in random donor DNA insertions, as determined by clonal analysis using junction PCR assays (117). Presumably, albeit more efficacious for generating populations of genetically modified cells, linear free-ended DNA is prone to homology-independent capture at chromosomal DSBs (targeted or otherwise) through illegitimate recombination processes comprising end-to-end DNA ligations.

Li and colleagues applied HC-AdV5/35 vectors for delivering into human $CD34^+$ cells *AAVS1*-specific RGNs and donor DNA templates encoding EGFP and the positive selectable marker $mgmt^{P140K}$ (118). The latter gene product confers resistance to O6BG/bis-chloroethylnitrosourea (BCNU). In this study, *AAVS1* gRNA target sites flanked the donor template for enhancing the frequencies of genetically modified cells via RGN-induced donor DNA excision. Co-transduction of human $CD34^+$ cells with both AdVs resulted in 0.9% of $EGFP^+$ hematopoietic cell clones as determined by CFU assays. Further characterization of these colonies (n=14) showed accurate insertion of the donor DNA at the *AAVS1* locus. The delivery of *AAVS1*-specific RGNs and *AAVS1*-targeting donor templates into murine Lin^- cells, isolated from the bone marrow of human *AAVS1*/*CD46* transgenic mice, was done through their ex vivo co-transduction with HC-AdV5/35 particles. As controls, parallel samples of Lin^- cells were exposed exclusively to one of the two vectors. Subsequently, vector-transduced Lin^- cells were transplanted into lethally irradiated C57BL/6 mice. Notably, in these experiments, no significant differences in engraftment rates were observed in mice receiving cells treated with the different HC-AdV5/35 regimens. At 4 weeks post-transplantation, an average of 1.1% and <0.2% of $EGFP^+$ peripheral blood mononuclear cells (PBMCs) were measured in the experimental and control groups, respectively. After three rounds of BCNU selection an enrichment in $EGFP^+$ cell marking was observed that varied from ~20 to ~100%, depending on the recipient mouse analyzed. Importantly, multilineage $EGFP^+$ cell marking was stably maintained for 16 weeks in secondary recipients demonstrating genetic modification of bona fide murine HSCs. Building on these data and experimental settings, Li and colleagues went on to test HDR-mediated knock-in of a γ-globin-coding transgene at the human *AAVS1* locus in murine Lin^- cells isolated from *AAVS1*/*CD46* transgenic mice. The transgene was placed under the regulation of a mini-β-globin locus control region for preferential expression in erythroid cells. Lin^- cells transduced with HC-AdV5/35 particles were transplanted into lethally irradiated C57BL/6 mice and were subsequently subjected to three rounds of BCNU selection. At 16 weeks post-transplantation, the level of γ-globin was on average 20.52% and 22.33% of that of adult mouse β-globin as measured by high-performance liquid chromatography and qRT-PCR analyses, respectively (118).

The cumulative data from these investigations on the use of AdV systems for gene editing of adult stem cells and their progeny bodes well for their application in basic research and biotechnologies, including for the development of genetic therapies targeting acquired and inherited disorders.

3. Human Embryonic Stem Cells (hESCs) and Human Induced Pluripotent Stem Cells (hiPSCs) Genome Editing

Human pluripotent stem cells (hPSCs) renown rose ever since the first isolation of human embryonic stem cells (hESCs) from pre-implantation embryos in 1998 (119). Under well-defined culture conditions,

hESCs are able to self-renew and can replicate for long periods in vitro while maintaining their full potential to differentiate into any somatic cell type derived from the three embryonic germ layers; endoderm, ectoderm, and mesoderm. These unique features of self-renewal and pluripotency facilitate studying cell differentiation processes and creating in vitro models of human disorders ("disease-in-a-dish"). In addition, hESCs hold the promise of revolutionizing regenerative medicine through the establishment of innovative stem cell therapies and represent invaluable tools for drug screening and development. Nevertheless, the therapeutic application of hESCs is limited not only by technical challenges but also ethical concerns stemming from their human-embryo origins (120). For this reason, the generation of human induced pluripotent stem cells (hiPSCs) represented a fundamental turning point in this field of biomedical research (121). This revolutionizing discovery took place in 2006, when Takahashi, Yamanaka and colleagues discovered that a cocktail of four transcription factors (i.e., KLF4, c-MYC, OCT4, and SOX2) was capable of reprogramming somatic, terminally differentiated cells, "back" to an hESC-like state (122,123). Indeed, for the most part, hiPSCs maintain the characteristics of hESCs, including their defining features of self-renewal and pluripotency. Crucially, cellular reprogramming overcomes the ethical concerns associated with hESCs and offers the possibility for generating and differentiating hiPSCs from virtually any individual into tissue-specific cell types. These capabilities permit in vitro disease modeling and drug screenings (124,125). Moreover, hiPSCs open the perspective for autologous cell transplantation therapies for repairing tissues and organs affected by injuries or, when combined with gene-editing technologies, inherited disorders (124,125) (**Figure 4**). Indeed, the advances made in gene editing technologies are greatly impacting hPSC-based research (126). Firstly, gene editing of hiPSCs is an important steppingstone towards their clinical translation, in that targeted correction of patient-derived hiPSCs might pave the way for the development of personalized regenerative medicines of otherwise untreatable genetic diseases (126) (**Figure 4**). Secondly, gene editing contributes to the establishment of clear genotype-phenotype associations by permitting the generation of isogenic pairs of hiPSC lines that share the same genetic background and differ exclusively in specific well-defined DNA sequences. These isogenic hiPSC pairs can be obtained either via correcting a genetic defect in a patient-derived hiPSC line or introducing mutations causing a genetic defect in a wild-type hiPSC line (**Figure 4**).

Several studies employing engineered ZFNs, TALENs, and RGNs, have shown the utility of these molecular tools for gene editing in hPSCs (127). The off-target effects and unpredictable genomic changes resulting from the repair of DSBs made by programmable nucleases are, however, major concerns in the gene editing field, especially in its application to stem cells (56-60). In this regard, recent developments on genome engineering strategies based on sequence- and strand-specific nucleases (nickases) as such (61,128-130) or on the fusion of these nickases to cytidine or adenine deaminases (i.e., base editors) (131) or reverse transcriptases (i.e., prime editors) (132) is gaining momentum. In part, this momentum derives from the fact that these tools open up the perspective for efficient, DSB-free, genetic modification of stem cells whose sensibility to DSBs is particularly acute (115,133,134). Next to gene editing strategies based on nucleases and nickases, there are also gene editing approaches that rely on the exclusive delivery of exogenous HDR substrates into hPSCs. In this case, stringent positive and negative selection schemes are often necessary for the isolation of properly targeted cells as HDR events are very rare in the absence of DSBs at target DNA (25-27) or SSBs at target and donor DNA (61,128-130,135). Moreover, to ameliorate the inefficiency of HDR in the absence of targeted DNA lesions, whenever possible, donor templates are endowed with long sequences homologous to target genomic regions. Indeed, extensive homologous sequences, normally spanning several thousands of bps flanking the desired exogenous DNA are exploited for obtaining site-specific gene insertion through spontaneous HDR. However, regardless of their dependency on or independency from nucleases or nickases, and derivatives thereof, a main challenge for operational gene editing in adult stem cells and hPSCs remains the need for delivering the necessary molecular tools in an efficient and, ideally, non-cytotoxic manner. To this end, various viral and non-viral delivery systems are being explored (9,136). We will next highlight the contributions of HC-AdV technology for gene editing in hiPSCs and hESCs

3.1. High-Capacity Adenoviral Vector (HC-AdV)-Based Gene Editing in hESCs and hiPSCs

HC-AdV-based gene editing of PSCs involving exclusively donor DNA delivery was initially applied in murine ESCs for achieving HDR-mediated correction of *Hprt* alleles (137). Soon thereafter, Suzuki and coworkers tested HC-AdVs for gene editing in hESCs (138). These authors started by comparing HC-

AdVs displaying serotype 5 or serotype 35 fibers for transducing hESCs by measuring through flow cytometry the frequencies of cells transiently expressing the Venus fluorescent protein reporter. Both viral vectors showed a clear multiplicity of infection (MOI)-dependent increase in transduction efficiencies that reached over 90% of target cells. The highest gene transfer levels were obtained with the tropism-modified vector. Notably, at a low to moderate MOI range, i.e., 10–300 transducing units per cell (TU/cell), cytotoxic effects were not significantly different from mock-transduced cells. Subsequently, HC-AdVs displaying conventional serotype 5 fibers were employed at a MOI of 300 TU/cell to deliver an *HRPT1*-targeting construct with long regions of homology (i.e., 14.3 kb and 9.2 kb) designed to insert a *neomycin phosphotransferase* (*neo*^{R}) cassette. Cells stably expressing the neoR gene product acquire resistance to the aminoglycoside antibiotic G418 (also known as geneticin) (**Figure 5**). In addition to the positive selection marker gene neoR, in order to minimize the expansion of cells with ectopic vector DNA integration, the vector genome also contained a negative selection cassette external to the homology regions expressing the Herpes Simplex Virus type 1 thymidine kinase (HSV1-tk) (**Figure 5**). Therefore, in case of HDR-independent or random chromosomal integration of HC-AdV DNA, stable HSV1-tk synthesis converts the pro-drug ganciclovir (GCV) into a phosphorylated cytotoxic product that leads to cell death (**Figure 5**). Among 5.1×10^{6} transduced hESCs, 136 colonies were G418-resistant and, of these, 31 were G418/GCV double-resistant. PCR and Southern blot analyses further demonstrated that of the 31 double-resistant colonies, 14 were correctly targeted at *HPRT1* (138). Importantly, HC-AdV transductions led to significantly higher gene transfer efficiencies than those obtained by "naked" DNA transfections based on electroporation and FuGENE HD. Moreover, when compared to the electroporation of the same *HPRT1*-targeting construct, HC-AdV donor delivery proved to be ~300 fold more efficient in terms of the frequencies of precisely edited cells obtained (138).

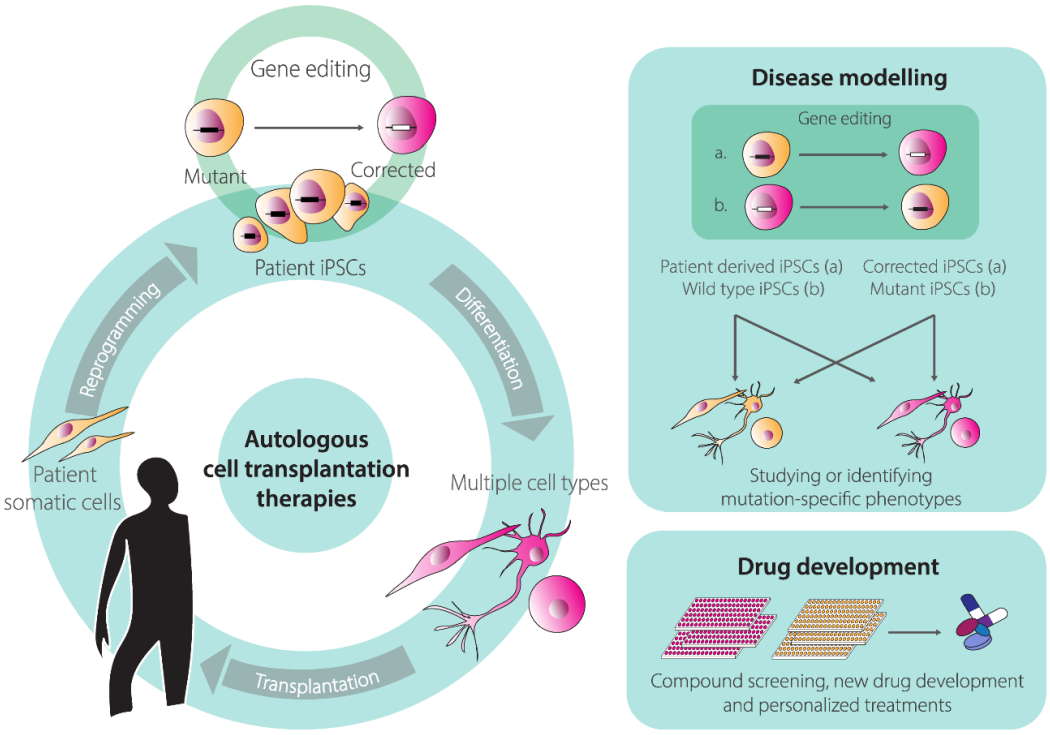


Figure 4. Illustration of human induced pluripotent stem cell (hiPSC)-based research and development activities enabled by genome editing technologies. Ex vivo reprogramming of patient-derived somatic cells into hiPSCs followed by their genetic correction, expansion, and directed differentiation into specialized cells types opens the perspective for the development of innovative autologous cell therapies. Generation of hiPSC lines sharing the same genetic background and differing from each other at predefined genetic loci can be accomplished via either (a) targeted correction of specific mutations in patient-derived hiPSCs or (b) targeted installation of specific mutations in wild-type, healthy donor-derived, hiPSCs. The resulting pairs of isogenic hiPSC lines form tractable experimental systems for the controlled and robust establishment of genotype-phenotype associations during disease modeling and for high-throughput screens aiming at assessing drug toxicities and/or identifying new drug candidates.

Building on these promising findings, a follow-up study investigated a similar HC-AdV-based gene editing approach in both hESCs and hiPSCs (139). In this study, the authors explored different gene

editing settings, i.e., (i) knock-in of a donor *neoR* cassette at the housekeeping *HPRT1* locus, (ii) knock-in of a donor *neoR* cassette designed for conditional knock-out of target genes located at different genomic positions, and (iii) knock-in of a donor *EGFP* cassette at a transcriptionally inactive *HB9* locus. Firstly, *HRPT1*-targeting experiments for knocking-in the donor *neoR* cassette in two distinct hiPSC lines led to 20% and 7% of correctly targeted clones after positive-negative G418/GCV selection (139). Significantly, control experiments involving the electroporation of the linearized *HPRT1*-targeting HC-AdV plasmid led to 0% of correctly targeted clones. Secondly, *neoR* cassette knock-in experiments at *KU80*, *LIG1*, and *LIG3* led to 81%, 34%, and 42% gene targeting frequencies, respectively. Subsequently, the loxP-flanked neoR cassette was excised in ~25% of the targeted cells through transient Cre delivery and target gene knockouts were confirmed through clonal analyses using Southern blotting, RT-qPCR, and western blotting. Finally, HC-AdV-mediated *EGFP* knock-in at the transcriptionally inactive HB9 locus led to 23% and 57% of accurate gene targeting in hiPSC and hESC lines, respectively. Other studies confirmed that silent loci are accessible to HC-AdV-based gene editing. For example, to trace gene expression during cell differentiation, HC-AdVs were employed to knock-in live-cell reporter genes into *ALB* and *OC* alleles to monitor the differentiation of hESCs and/or hiPSCs along the hepatic and osteogenic lineages, respectively (140,141).

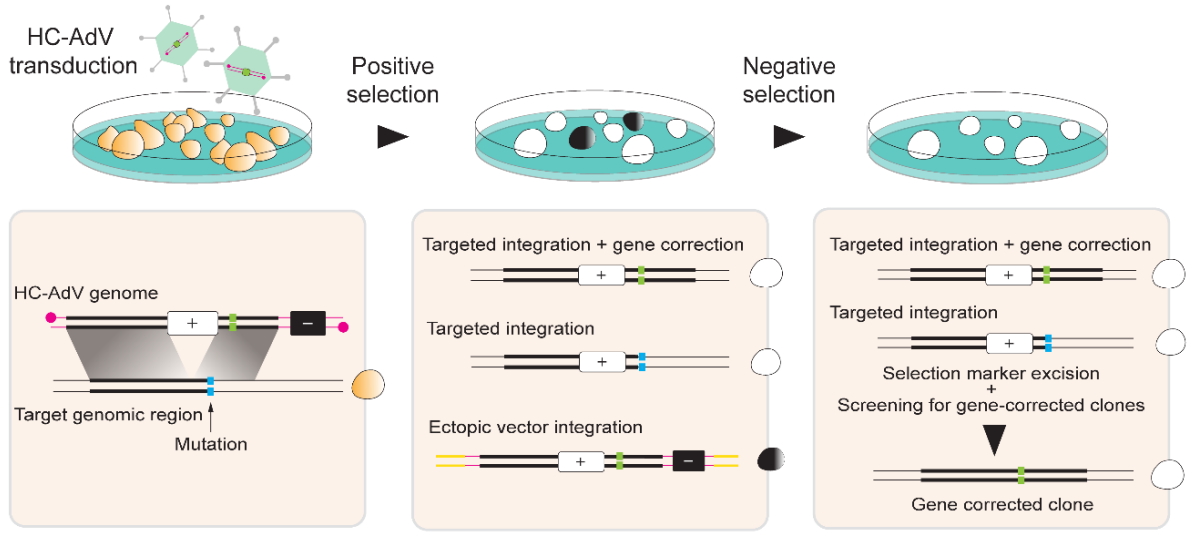


Figure 5. DSB-independent gene editing based on HC-AdV donor DNA transduction and positive-negative cell selection protocols. HC-AdV genomes contain a positive selection cassette, e.g., *neo*^{R} (white box) flanked by extensive human DNA sequences that are homologous to a target genomic region except for specific nucleotide(s) (left panel). In this example, donor and acceptor templates bear wild-type and mutant allelic sequences of a target gene (green and cyan boxes, respectively) so that, after recombination, involving outward homologous regions, gene correction ensues (middle panel, upper diagram). Next to these wanted outcomes there are also unwanted ones in the form of homologous and non-homologous recombination events resulting in no gene correction and random chromosomal donor DNA integration (middle panel, central, and bottom diagrams, respectively). Cells containing these different types of genetic modifications survive and multiply in the presence of a cell-killing drug that is broken-down by the positive-selection gene product. Selective elimination of cells with random HC-AdV donor DNA insertions is accomplished owing to the presence of a suicide negative selection cassette located outside the homology regions, e.g., *HSV-tk* (black box), that convers a prodrug substrate into a cell-killing product. The positive selection marker can subsequently be removed by site-specific recombinases, e.g., Cre and FLP that leave *lox*P and *FRT* site footprints, respectively, in the genome. Alternatively, transposon/transposase systems, e.g., footprint-free PiggyBac variants can be used that ultimately achieve scarless genomic modifications. Finally, genotyping screens permit identifying cells containing correctly targeted alleles (right panel).

3.2. HC-AdV-Based Gene Editing for Targeted Gene Correction in Human Pluripotent Stem Cells (hPSCs)

HC-AdVs are also being investigated for targeted correction of disease-causing mutations in hPSCs (**Figure 5**). Initial experiments targeted mutations underlying Hutchinson–Gilford progeria syndrome (HGPS) and atypical Werner syndrome (AWS) in hiPSCs (142). HGPS and AWS are laminopathies whose mutations in the exon 11 of the *LMNA* gene include C1824T and A1733T, respectively. These mutations affect the nuclear structure resulting in premature aging. By exploiting the large cloning capacity of HC-AdV particles, HGPS and AWS can potentially be tackled by a single large *LMNA*-targeting construct covering different mutations. Similar to previous work (138), upon HC-AdV donor DNA transduction of hiPSCs and positive-negative G418/GCV selection, integration of the *neoR* cassette

at the *LMNA* target site between exons 10 and 11 ranged from 78% to 100%, as assessed through PCR and Southern blot analyses (142). Correction of the 1-bp substitutions C1824T and A1733T located in exon 11 of *LMNA* in HGPS-hiPSCs and AWS-hiPSCs, respectively, was verified through DNA sequencing of targeted clones. This analysis revealed that 12 out of 25 HGPS-hiPSC clones and 35 out of 65 AWS-hiPSCs clones were accurately repaired. Subsequently, the *neoR* cassette, flanked by FRT sites, was excised by transient expression of FLPe recombinase leading to wild-type *LMNA* expression and subsequent rescue of the HGPS phenotype, as determined by the restoration of normal nuclear architecture and cell senescence programs (142). Next, in addition to confirming the pluripotency of gene-edited hiPSCs, the authors meticulously investigated the genetic and epigenetic integrity of the corrected cells. In particular, correctly targeted hiPSCs showed a normal karyotype, expressed pluripotency markers and exhibited demethylation of the promoter of the pluripotency gene *OCT4* (142). Moreover, genome-wide single nucleotide polymorphism (SNP), DNA microarray, and genome-wide DNA methylation analyses indicated a generic maintenance of the genetic background, global gene expression patterns, and global epigenetic states, respectively, in gene-edited cells using parental hiPSC lines as references (142). In another study, HC-AdV-based gene editing was applied to correct the A→T transversion at nucleotide 20 in exon 1 of the β-globin-encoding *HBB* gene in hiPSCs obtained from SCD patients (143). In these experiments, the positive-negative G418/GCV selection resulted in an average of 85% of colonies with *neoR* targeted insertions with an average of 81% of these colonies presenting the desired HBB gene correction (143).

The previously described gene editing experiments targeting *LMNA* (142) and *HBB* (143), demonstrated that vector DNA-derived SNPs could be found in the correctly targeted clones at positions 4.4-kb and 3.6-kb away from the *neoR* insertion site within *LMNA* and *HBB* alleles, respectively. On the basis of these results, the authors postulated that the HC-AdV platform might be valuable for repairing mutations found in a relatively broad target region, increasing its potential as a versatile gene correction tool. As an example, a single *LMNA*-targeting HC-AdV could potentially repair over 200 *LMNA* mutations associated with laminopathies (142).

Two subsequent studies sought to formally investigate; (i) the extent of homology between endogenous target and exogenous HC-AdV donor templates required for efficient gene editing (144); and (ii) the relationship between the distance from the knock-in target site and the incorporation of polymorphic markers located along the region of homology (145). In both studies, HC-AdV targeting constructs were directed to the *CFTR* locus in a hiPSC line harboring the heterozygous mutations ΔF508 and ΔI507 in exon 10 of the target gene. To investigate the effect of the extent of homology on the efficiency of HC-AdV-based gene editing, a set of five different HC-AdVs containing differently sized wild-type *CTFR* sequences were tested (144). The homology regions spanned total lengths of 23.8 kb, 21.4 kb, 14.8 kb, 9.6 kb, and 5.6 kb. Transduction of hiPSCs with the various HC-AdV donors followed by G418 and GCV double selection led to the emergence of colonies that were subsequently subjected to Southern blot analysis for determining the frequencies of targeted events. The HC-AdV donor construct carrying 23.8 kb of sequence homology to genomic DNA led to 97.4–100% of gene-targeted clones; whilst the HC-AdV donor construct bearing 5.6 kb of sequence homology to genomic DNA yielded 50% of gene-targeted clones (144). Together, these data lend additional support to a direct correlation between the length of homology between target and donor DNA and the frequency of HDR-mediated gene targeting (21).

In order to investigate the extent of exchange of homologous sequences between target and donor DNA templates, twelve 2-bp insertions were introduced along the 23.8 kb homology region in a *CTFR*-targeting HC-AdV construct (145). Upon HC-AdV-mediated gene targeting, each of these 2-bp insertions convert an endogenous restriction enzyme recognition site into that of another allowing for straightforward assessment of the extent of recombination between target and donor DNA sequences. As assessed through Southern blot analysis, 89.5% of drug-resistant hiPSC clones were correctly targeted at *CFTR* alleles (145). Furthermore, PCR and restriction enzyme fragment length analyses of the drug-selected hiPSC clones showed that the closest marker to the insertion site (i.e., 208 bp) was incorporated in 100% of the analyzed clones. Conversely, the most distant marker to the insertion site (i.e., 11.2 kb) was incorporated in only 21.7% of the analyzed clones, suggesting that the vicinity of polymorphic markers to the insertion site is proportional to their genomic incorporation rate. Interestingly, 4.8% of the clones presented all the twelve restriction enzyme markers. This data suggests

that HC-AdV-based gene editing can be used to introduce genetic information distributed over a wide range of homologous DNA in hiPSCs (i.e., at least up to 22.2 kb) (145). As aforementioned, HC-AdV-based gene editing is equally applicable for establishing tractable in vitro disease models comprising pairs of isogenic hPSC lines whose genomes differ at well-defined locations (**Figure 4**). Indeed, HC-AdV-based gene editing has been explored for modeling various human disorders, including; Parkinson's disease (146), Fanconi anemia (147), retinitis pigmentosa (148), and Werner syndrome (149). Combining HC-AdV and programmable nuclease technologies offers the prospect for improving gene editing frequencies. In this regard, Suzuki and colleagues used HC-AdVs to deliver donor templates alone or together with TALEN expression units (150). The TALEN and donor HDR substrates were tailored for targeting *HBB* alleles underlying SCD in hiPSC lines. Transduction of SCD patient-derived hiPSCs with the "all-in-one" HC-AdV resulted in an increase in gene-targeting frequencies when compared to those achieved by HC-AdV delivery of donor DNA templates alone (150). Specifically, among 2×10^5 cells transduced with the "all-in-one" HC-AdV, 28 G418-resistant clones were analyzed and of these 86% were correctly targeted. Conversely, among 9×10^6 cells transduced with an HC-AdV delivering exclusively donor DNA, 134 G418-resistant clones were analyzed with only 22% of these being correctly targeted (150).

The cumulative data on HC-AdV-based gene editing in hPSCs bodes well for its application in basic research, drug screening, disease modeling, and eventually, development of autologous cell therapies for inherited disorders (**Figure 4**).

4. Conclusions and Outlook

Rapid advancements in the gene editing and stem cell fields are contributing to broaden the range of options for addressing scientific questions and developing candidate gene and cell therapies. To support the integration of these fields, and hence further widen their reach, it is crucial to develop delivery systems that permit introducing programmable DNA-targeting enzymes and donor nucleic acid templates into target cells in an efficient and versatile manner. Moreover, additional parameters that need to be taken into consideration concern the effects that the delivery systems themselves have on the ultimate performance and accuracy of gene editing procedures. In the case of gene-editing tool delivery through viral vector systems, it is important that vector genomes transporting donor templates or encoding programmable DNA-targeting enzymes are refractory to (i) structural rearrangements (103), (ii) epigenetic silencing mechanisms (106,107), and (iii) capture at chromosomal DSBs via illegitimate recombination processes (117,151,152).

Recent developments on genomic engineering comprise the progression from chromosomal cutting to chromosomal non-cutting approaches based on nicking Cas9 variants and on these variants fused to heterologous DNA-modifying moieties. These new gene editing principles include; (i) HDR-mediated chromosomal insertion of exogenous DNA spanning from single bps to whole transgenes through SSB formation at target and donor DNA (61,128-130), and (ii) donor DNA-free in situ installation of genetic changes through base editing (131) or prime editing (132). Base editors, comprising a Cas9 nickase covalently linked to a cytidine or adenine deaminase, induce C→T or A→G transitions, respectively (153,154). These conversions occur within so-called "editing windows" located in target sequences defined by a standard gRNA (131,153,154). Prime editors, consisting of a Cas9 nickase covalently linked to an engineered oncoretroviral reverse transcriptase (RT), in addition to transitions, also generate defined indels and transversions, e.g., A→C, G→T, T→A, and C→G (132). The exact genetic modification depends on the designing of an extended gRNA dubbed prime editor gRNA (pegRNA). The pegRNA is formed by the standard gRNA sequences crRNA and tracrRNA (Figure 1C) covalently linked to a RT primer binding site (PBS) and a RT template sequence bearing the intended edit. After nicking, the PBS locally anneals to the 3′-ended DNA flap that primes RT synthesis over the RT template. The resulting DNA copy of the edit ultimately becomes incorporated at the genomic target site upon a series of cellular processing steps responsible for removing DNA flaps that do not hybridize to target sequences (132).

The SSB-mediated gene editing approaches are opening the perspective for modifying complex genomes with unprecedented precision while minimizing unwanted events characteristic of DSB-mediated gene editing procedures. In addition to off-target mutagenesis (56-61), unwanted genome-modifying events include translocations (60,61) and unpredictable genomic "scars" at target sequences

in the form of indels and larger structural rearrangements resulting from site-specific DSB repair via prevalent NHEJ pathways (60,155). Not surprisingly, however, new gene editing approaches and technologies bring to the fore their own sets of shortcomings that need to be carefully assessed and resolved. For instance, base editors can yield off-target editing at the genome and transcriptome levels (156); whereas primer editing can install target-site mutations derived from RT synthesis into the pegRNA scaffold (132). Although the optimization of gene editing tools and strategies should ideally take place in the target cell types of interest, each of which bearing its specific epigenome, these investigations are rendered difficult due to the fact that latest-generation gene editing tools are becoming even larger than the original Cas9:gRNA complexes. Indeed, prime editors and base editors consist of a bulky Cas9 nickase fused to one and two, respectively, heterologous proteins that must work together as large macromolecular machines (67,131,132,153,154). Therefore, there is a pressing need for developing and testing delivery vehicles that can introduce such large machines into primary human cells so that their performance and interaction with human (epi)genomes can be thoroughly investigated. In this context, the research reviewed herein on the testing and use of AdV systems for the targeted genetic modification of stem cells, progenitor cells, and their progeny, supports the view that these agents will become increasingly applied for achieving flexible gene-editing tool delivery and precise gene-editing outcomes in human cells. Defining features underpinning the suitability of AdVs for investigating new gene-editing modalities include their efficient transduction of cycling and quiescent cells, amenability to tropism modifications, high genetic stability and strict episomal nature. Moreover, in the case of HC-AdVs, the absence of viral genes and vast packaging capacity (i.e., up to 36 kb) makes this platform particularly suited for ferrying into cells large genetic payloads for testing precision gene-editing principles based on the recruitment of the HDR pathway or the delivery of DNA-editing fusion constructs, e.g., base and prime editors.

ACKNOWLEDGMENTS

The authors thank Prof. Rob Hoeben (Department of Cell and Chemical Biology, LUMC, The Netherlands) for providing the transmission electron micrograph of the adenovirus serotype 5 virion and for critically reading the manuscript. The authors also thank Marcella Dias Brescia and Hidde Zittersteijn (both from the Department of Cell and Chemical Biology, LUMC, The Netherlands) for critically reading the manuscript.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 765269 (IMGENE – Improving Genome Editing Efficiency). The research in our laboratory is further supported by the Dutch Prinses Beatrix Spierfonds (W.OR11-18) and the Duchenne Parent Project NL (17.012). Q.W. is the recipient of a Ph.D. fellowship from the China Scholarship Council-Leiden University Joint Scholarship Program.
Conflict of interest statement. None declared.

REFERENCES

1. Chang, H.H.Y.; Pannunzio, N.R.; Adachi, N.; Lieber, M.R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell. Biol*. **2017**, *18*, 495–506.
2. Kim, Y.; Kweon, J.; Kim, J.S. TALENs and ZFNs are associated with different mutation signatures. *Nat. Methods* **2013**, *10*, 185.
3. Allen, F.; Crepaldi, L.; Alsinet, C.; Strong, A.J.; Kleshchevnikov, V.; De Angeli, P.; Páleníková, P.; Khodak, A.; Kiselev, V.; Kosicki, M.; et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol*. **2019**, *37*, 64–72.
4. Ata, H.; Ekstrom, T.L.; Martínez-Gálvez, G.; Mann, C.M.; Dvornikov, A.V.; Schaefbauer, K.J.; Ma, A.C.; Dobbs, D.; Clark, K.J.; Ekker, S.C. Robust activation of microhomology-mediated end joining for precision gene editing applications. *PLoS Genet*. **2018**, *14*, e1007652.
5. Taheri-Ghahfarokhi, A.; Taylor, B.J.M.; Nitsch, R.; Lundin, A.; Cavallo, A.L.; Madeyski-Bengtson, K.; Karlsson, F.; Clausen, M.; Hicks, R.; Mayr, L.M.; et al. Decoding non-random mutational signatures at Cas9 targeted sites. *Nucleic Acids Res*. **2018**, *46*, 8417–8434.
6. Shen, M.W.; Arbab, M.; Hsu, J.Y.; Worstell, D.; Culbertson, S.J.; Krabbe, O.; Cassa, C.A.; Liu, D.R.; Gifford, D.K.; Sherwood, R.I. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **2018**, *563*, 646–651.

7. van Overbeek, M.; Capurso, D.; Carter, M.M.; Thompson, M.S.; Frias, E.; Russ, C.; Reece-Hoyes, J.S.; Nye, C.; Gradia, S.; Vidal, B.; et al. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **2016**, *63*, 633–646.
8. Ousterout, D.G.; Perez-Pinera, P.; Thakore, P.I.; Kabadi, A.M.; Brown, M.T.; Qin, X., Fedrigo, O.; Mouly, V.; Tremblay, J.P.; Gersbach, C.A. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Mol. Ther*. **2013**, *21*, 1718–1726.
9. Chen, X.; Gonçalves, M.A. Engineered Viruses as Genome Editing Devices. *Mol. Ther*. **2016**, *24*, 447–457.
10. Doudna, J.A.; Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096.
11. Chandrasegaran, S.; Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J. Mol. Biol*. **2016**, *428*, 963–989.
12. El-Brolosy, M.A.; Kontarakis, Z.; Rossi, A.; Kuenne, C.; Günther, S.; Fukuda, N.; Kikhi, K.; Boezio, G.L.M.; Takacs, C.M.; et al. Genetic compensation triggered by mutant mRNA degradation. *Nature* **2019**, *568*, 193–197.
13. Ma, Z.; Zhu, P.; Shi, H.; Guo, L.; Zhang, Q.; Chen, Y.; Chen, S.; Zhang, Z.; Peng, J.; Chen, J. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* **2019**, *568*, 259–263.
14. Tuladhar, R.; Yeu, Y.; Piazza, J.T.; Tan, Z.; Clemenceau, J.R.; Wu, X.; Barrett, Q.; Herbert, J.; Mathews, D.H.; Kim, J.; et al. CRISPR–Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat. Commun*. **2019**; *10*, 4056.
15. Söllü, C.; Pars, K.; Cornu, T.I.; Thibodeau-Beganny, S.; Maeder, M.L.; Joung, J.K.; Heilbronn, R.; Cathomen, T. Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res*. **2010**, *38*, 8269–8276.
16. Ousterout, D.G.; Kabadi, A.M.; Thakore, P.I.; Majoros, W.H.; Reddy, T.E.; Gersbach, C.A. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun*. **2015**, *6*, 6244.
17. Maggio, I.; Stefanucci, L.; Janssen, J.M.; Liu, J., Chen, X.; Mouly, V.; Gonçalves, 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.
18. Choi, P.S.; Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun*. **2014**, *24*, 5:3728.
19. Yamamoto, Y.; Gerbi, S.A. Making ends meet: Targeted integration of DNA fragments by genome editing. *Chromosoma* **2018**, *127*, 405–420.
20. Scully, R.; Panday, A.; Elango, R.; Willis, N.A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell. Biol.* **2019**, *20*, 698–714.
21. Deng, C.; Capecchi, M.R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol*. **1992**, *12*, 3365–3371.
22. Porter, A.C.; Itzhaki, J.E. Gene targeting in human somatic cells. Complete inactivation of an interferon-inducible gene. *Eur. J. Biochem*. **1993**, *218*, 273–281.
23. Ganguly, A.; Smelt, S.; Mewar, R.; Fertala, A.; Sieron, A.L.; Overhauser, J.; Prockop, D.J. Targeted insertions of two exogenous collagen genes into both alleles of their endogenous loci in cultures human cells: The insertions are directed by relatively short fragments containing the promoters and the 5′ ends of the genes. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7365–7369.
24. Thyagarajan, B.; Johnson, B.L.; Campbell, C. The effect of target site transcription on gene targeting in human cells in vitro. *Nucleic Acids Res*. **1995**, *23*, 2784–2790.
25. Lukacsovich, T.; Yang, D.; Waldman, A.S. Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. *Nucleic Acids Res*. **1994**, *22*, 5649–5657.
26. Rouet, P.; Smih, F.; Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6064–6068.
27. Rouet, P.; Smih, F.; Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol*. **1994**, *14*, 8096–8106.
28. Maggio, I.; Gonçalves, M.A. Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends Biotechnol*. **2015**, *33*, 280–291.

29. Kim, Y.G.; Cha, J.; Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1156–1160.
30. Christian, M.; Cermak, T.; Doyle, E.L.; Schmidt, C.; Zhang, F.; Hummel, A.; Bogdanove, A.J.; Voytas, D.F. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **2010**, *186*, 757–761.
31. Li, T.; Huang, S.; Zhao, X.; Wright, D.A.; Carpenter, S.; Spalding, M.H.; Weeks, D.P.; Yang, B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*. **2011**, *39*, 6315–6325.
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. USA* **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. Biotechnol*. **2011**, *29*, 143–148.
34. Mussolino, C.; Morbitzer, R.; Lütge, F.; Dannemann, N.; Lahaye, T.; Cathomen, T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*. **2011**, *39*, 9283–9293.
35. Cho, S.W.; Kim, S.; Kim, J.M.; Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol*. **2013**, *31*, 230–232.
36. Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, *339*, 819–823.
37. Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-programmed genome editing in human cells. *Elife* **2013**, *2*, e00471.
38. 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.
39. Miller, J.; McLachlan, A.D.; Klug, A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J*. **1985**, *4*, 1609–1614.
40. Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **2009**, *326*, 1509–1512.
41. Moscou, M.J.; Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **2009**, *326*, 1501.
42. 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.
43. 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.
44. Bitinaite, J.; Wah, D.A.; Aggarwal, A.K.; Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10570–10575.
45. Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem*. **2010**, *79*, 213–231.
46. Ecco, G.; Imbeault, M.; Trono, D. KRAB zinc finger proteins. *Development* **2017**, *144*, 2719–2729.
47. Boch, J.; Bonas, U. Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol*. **2010**, *48*, 419–436.
48. Cathomen, T.; Joung, K.J. Zinc-finger Nucleases: The next generation emerges. *Mol. Ther*. **2008**, *16*, 1200–1207.
49. Mussolino, C.; Cathomen, T. TALE nucleases: Tailored genome engineering made easy. *Curr. Opin. Biotechnol*. **2012**, *23*, 644–650.
50. Bultmann, S.; Morbitzer, R.; Schmidt, C.S.; Thanisch, K.; Spada, F.; Elsaesser, J.; Lahaye, T.; Leonhardt, H. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res*. **2012**, *40*, 5368–5377.
51. 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.

52. Chen, X.; Rinsma, M.; Janssen, J.M.; Liu, J.; Maggio, I.; Gonçalves MA. Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res*. **2016**, *44*, 6482–6492.
53. Makarova, K.S.; Wolf, Y.I.; Iranzo, J.; Shmakov, S.A.; Alkhnbashi, O.S.; Brouns, S.J.J.; Charpentier, E.; Cheng, D.; Haft, D.H.; Horvath, P.; et al. Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol*. **2020**, *18*, 67–83.
54. 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.
55. Shibata, M.; Nishimasu, H.; Kodera, N.; Hirano, S.; Ando, T.; Uchihashi, T.; Nureki, O. Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat. Commun*. **2017**, *8*, 1430.
56. Cradick, T.J.; Fine, E.J.; Antico, C.J.; Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*. **2013**, *41*, 9584–9592.
57. Hsu, P.D.; Scott, D.A.; Weinstein, J.A.; Ran, F.A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E.J.; Wu, X.; Shalem, O.; et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol*. **2013**, *31*, 827–832.
58. Fu, Y.; Foden, J.A.; Khayter, C.; Maeder, M.L.; Reyon, D.; Joung, J.K.; Sander, J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol*. **2013**, *31*, 822–826.
59. Lin, Y.; Cradick, T.J.; Brown, M.T.; Deshmukh, H.; Ranjan, P.; Sarode, N.; Wile, B.M.; Vertino, P.M.; Stewart, F.J.; Bao, G. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res*. **2014**, *42*, 7473–7485.
60. Frock, R.L.; Hu, J.; Meyers, R.M.; Ho, Y.J.; Kii, E.; Alt, F.W. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol*. **2015**, *33*, 179–186.
61. Chen, X.; Tasca, F.; Wang, Q.; Liu, J.; Janssen, J.M.; Brescia, M.D.; Bellin, M.; Szuhai, K.; Kenrick, J.; Frock, R.L.; et al. Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking. *Nucleic Acids Res*., **2020**, *48*, 974–995.
62. Cho, S.W.; Kim, S.; Kim, Y.; Kweon, J.; Kim, H.S.; Bae, S.; Kim, J.S. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. **2014**, *24*, 132–141.
63. Akcakaya, P.; Bobbin, M.L.; Guo, J.A.; Malagon-Lopez, J.; Clement, K.; Garcia, S.P.; Fellows, M.D.; Porritt, M.J.; Firth, M.A.; Carreras, A.; et al. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature*, **2018**, *561*, 416–419.
64. Chen, X., Liu, J., Janssen, J.M., Gonçalves, M.A. The chromatin structure differentially impacts high-specificity CRISPR-Cas9 nuclease strategies. *Mol. Ther. Nucleic Acids*, **2017**, *8*, 558–563.
65. 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.
66. Jensen, K.T.; Fløe, L.; Petersen, T.S.; Huang, J.; Xu, F.; Bolund, L.; Luo, Y.; Lin, L. Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett*., **2017**, *591*, 1892–1901.
67. Chen, X.; Gonçalves, M.A. DNA, RNA, and Protein Tools for Editing the Genetic Information in Human Cells. *iScience* **2018**, *6*, 247–263.
68. Brooks, A.K.; Gaj, T. Innovations in CRISPR technology. *Curr. Opin. Biotechnol*. **2018**, *52*, 95–101.
69. Hajizadeh Dastjerdi, A.; Newman, A.; Burgio, G. The Expanding Class 2 CRISPR Toolbox: Diversity, Applicability, and Targeting Drawbacks. *BioDrugs* **2019**, *33*, 503–513.
70. Wold, W.S.M.; Horwitz, M.S. Adenoviruses. *Fields Virology* (5th ed.), **2007**, Lippincott Williams & Wilkins, Philadelphia, PA.
71. Gonçalves, M.A.; de Vries, A.A. Adenovirus: From foe to friend. *Rev. Med. Virol*. **2006**, *16*, 167–186.
72. Russell, W.C. Adenoviruses: Update on structure and function. *J. Gen. Virol*. **2009**, *90*, 1–20.
73. Vellinga, J.; van der Heijdt, S.; Hoeben, R.C. The adenovirus capsid: Major progress in minor proteins. *J. Gen. Virol*. **2005**, *86*, 1581–1588.
74. Bergelson, J.M.; Cunningham, J.A.; Droguett, G.; Kurt-Jones, E.A.; Krithivas, A.; Hong, J.S.; Horwitz, M.S.; Crowell, R.L.; Finberg, R.W. Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **1997**, *275*, 1320–1323.

75. Tomko, R.P.; Xu, R.; Philipson, L. HCAR and MCAR: The human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3352–3356.
76. Gaggar, A.; Shayakhmetov, D.M.; Lieber, A. CD46 is a cellular receptor for group B adenoviruses. *Nat. Med*. **2003**, *9*, 1408–1412.
77. Wang, H.; Li, Z.Y.; Liu, Y.; Persson, J.; Beyer, I.; Möller, T.; Koyuncu, D.; Drescher, M.R.; Strauss, R.; Zhang, X.B.; et al. Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11 and 14. *Nat. Med*. **2011**, *17*, 96–104.
78. Nilsson, E.C.; Storm, R.J.; Bauer, J.; Johansson, S.M.; Lookene, A.; Ångström, J.; Hedenström, M.; Eriksson, T.L.; Frängsmyr, L.; Rinaldi, S.; et al. The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. *Nat. Med*. **2011**, *17*, 105–109.
79. Arnberg, N., Edlund, K., Kidd, A.H., Wadell, G. Adenovirus type 37 uses sialic acid as a cellular receptor. *J. Virol*. **2000**, *74*, 42–48.
80. Alonso-Padilla, J.; Papp, T.; Kaján, G.L.; Benkő, M.; Havenga, M.; Lemckert, A.; Harrach, B.; Baker, A.H. Development of Novel Adenoviral Vectors to Overcome Challenges Observed With HAdV-5-based Constructs. *Mol. Ther*. **2016**, *24*, 6–16.
81. Barry, M.A.; Rubin, J.D.; Lu, S.C. Retargeting adenoviruses for therapeutic applications and vaccines. *FEBS Lett*. **2020**. doi: 10.1002/1873-3468.13731.
82. Shayakhmetov, D.M.; Papayannopoulou, T.; Stamatoyannopoulos, G.; Lieber, A. Efficient gene transfer into human CD34+ cells by a retargeted adenovirus vector. *J. Virol*., **2000**, *74*, 2567–2583.
83. Knaän-Shanzer, S.; Van Der Velde, I.; Havenga, M.J.E.; Lemckert, A.A.C.; De Vries, A.A.F.; Valerio, D. Highly efficient targeted transduction of undifferentiated human hematopoietic cells by adenoviral vectors displaying fiber knobs of subgroup B. *Hum. Gene Ther*., **2001**, *12*, 1989–2005.
84. Gonçalves, M.A.; de Vries, A.A.; Holkers, M.; van de Watering, M.J.M.; van der Velde, I.; van Nierop, G.P.; Valerio, D.; Knaän-Shanzer, S. Human mesenchymal stem cells ectopically expressing full-length dystrophin can complement Duchenne muscular dystrophy myotubes by cell fusion. *Hum. Mol. Genet*., **2005**, *15*, 213–221.
85. Knaän-Shanzer, S.; van de Watering, M.J.M.; van der Velde, I.; Gonçalves, M.A.F.V.; Valerio, D.; de Vries, A.A.F. Endowing human adenovirus serotype 5 vectors with fiber domains of species B greatly enhances gene transfer into human mesenchymal stem cells. *Stem Cells* **2005**, *23*, 1598–1607.
86. Gonçalves, M.A.; Holkers, M.; Cudré-Mauroux, C.; Van Nierop, G.P.; Knaän-Shanzer, S.; Van Der Velde, I.; Valerio, D.; De Vries, A.A. Transduction of myogenic cells by retargeted dual high-capacity hybrid viral vectors: Robust dystrophin synthesis in duchenne muscular dystrophy muscle cells. *Mol. Ther*., **2006**, *13*, 976–986.
87. Meier, O.; Greber, U.F. Adenovirus endocytosis. *J. Gene Med*. **2004**, *6*, S152–S163.
88. Graham, F.L.; Smiley, J.; Russell,W.C.; Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol*. **1977**, *36*, 59–74.
89. 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.
90. Shirley, J.L.; de Jong, Y.P.; Terhorst, C.; Herzog, R.W. Immune Responses to Viral Gene Therapy Vectors. *Mol. Ther*. **2020**, *28*, 709–722.
91. Shukarev, G.; Callendret, B.; Luhn, K.; Douoguih, M. A two-dose heterologous prime-boost vaccine regimen eliciting sustained immune responses to Ebola Zaire could support a preventive strategy for future outbreaks. *Hum. Vaccin. Immunother*. **2017**, *13*, 266–270.
92. Barouch, D.H.; Tomaka, F.L.; Wegmann, F.; Stieh, D.J.; Alter, G.., Robb, M.L.; Michael, N.L.; Peter, L.; Nkolola, J.P.; Borducchi, E.N.; et al. 2018. Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). *Lancet* **2018**, *392*, 232–243.
93. Havenga, M.J.; Holterman, L.; Melis, I.; Smits, S.; Kaspers, J.; Heemskerk, E.; van der Vlugt, R.; Koldijk, M.; Schouten, G.J.; Hateboer, G.; et al. Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: High yield and preserved bioactivity. *Biotechnol. Bioeng*. **2008**, *100*, 273–283.

94. Parks, R.J.; Graham, F.L. A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. *J. Virol*. **1997**, *71*, 3293–3298.
95. Jager, L.; Hausl, M.A.; Rauschhuber, C.; Wolf, N.M.; Kay, M.A.; Ehrhardt, A. A rapid protocol for construction and production of high-capacity adenoviral vectors. *Nat. Protoc*. **2009**, *4*, 547–564.
96. Perez, E.E.; Wang, J.; Miller, J.C.; Jouvenot, Y.; Kim, K.A.; Liu, O.; Wang, N.; Lee, G.; Bartsevich, V.V.; Lee, Y.L.; et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **2008**, *26*, 808–816.
97. Li, L.; Krymskaya, L.; Wang, J.; Henley, J.; Rao, A.; Cao, L.-F.; Tran, C.-A.; Torres-Coronado, M.; Gardner, A.; Gonzalez, N.; et al. Genomic Editing of the HIV-1 Coreceptor CCR5 in Adult Hematopoietic Stem and Progenitor Cells Using Zinc Finger Nucleases. *Mol. Ther.* **2013**, *21*, 1259–1269.
98. Maier, D.A.; Brennan, A.L.; Jiang, S.; Binder-Scholl, G.K.; Lee, G.; Plesa, G.; Zheng, Z.; Cotte, J.; Carpenito, C.; Wood, T.; et al. Efficient clinical scale gene modification via zinc finger nuclease-targeted disruption of the HIV co-receptor CCR5. *Hum. Gene Ther.* **2013**, *24*, 245–258.
99. Saydaminova, K.; Ye, X.; Wang, H.; Richter, M.; Ho, M.; Chen, H.; Xu, N.; Kim, J.S.; Papapetrou, E.; Holmes, M.C.; et al. 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.* **2015**, *1*, 14057.
100. Tebas, P.; Stein, D.; Tang, W.W.; Frank, I.; Wang, S.Q.; Lee, G.; Spratt, S.K.; Surosky, R.T.; Giedlin, M.A.; Nichol, G.; et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **2014**, *370*, 901–910.
101. Provasi, E.; Genovese, P.; Lombardo, A.; Magnani, Z.; Liu, P.Q.; Reik, A.; Chu, V.; Paschon, D.E.; Zhang, L.; Kuball, J.; et al. Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat. Med.* **2012**, *18*, 807–815.
102. Yuan, J.; Wang, J.; Crain, K.; Fearns, C.; Kim, K.A.; Hua, K.L.; Gregory, P.D.; Holmes, M.C.; Torbett, B.E. Zinc-finger Nuclease Editing of Human cxcr4 Promotes HIV-1 CD4+ T Cell Resistance and Enrichment. *Mol. Ther.* **2012**, *20*, 849–859.
103. Holkers, M.; Maggio, I.; Liu, J.; Janssen, J.M.; Miselli, F.; Mussolino, C.; Recchia, A.; Cathomen, T.; Gonçalves, M.A. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* **2013**, *41*, e63.
104. 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.
105. Fang, Y.; Stroukov, W.; Cathomen, T.; Mussolino, C. Chimerization Enables Gene Synthesis and Lentiviral Delivery of Customizable TALE-Based Effectors. *Int. J. Mol. Sci.* **2020**, *21*, 795.
106. Pelascini, L.P.; Maggio, I.; Liu, J.; Holkers, M.; Cathomen, T.; Gonçalves, M.A. Histone deacetylase inhibition rescues gene knockout levels achieved with integrase-defective lentiviral vectors encoding zinc-finger nucleases. *Hum. Gene Ther. Methods* **2013**, *24*, 399–411.
107. Pelascini, L.P.; Janssen, J.M.; Gonçalves, M.A. Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. *Hum. Gene Ther.* **2013**, *24*, 78–96.
108. Maggio, I.; Holkers, M.; Liu, J.; Janssen, J.M.; Chen, X.; Gonçalves, M.A. 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.
109. Gong, H.; Liu, M.; Klomp, J.; Merrill, B.J.; Rehman, J.; Malik, A.B. Method for Dual Viral Vector Mediated CRISPR-Cas9 Gene Disruption in Primary Human Endothelial Cells. *Sci. Rep.* **2017**, *7*, 42127.
110. Maggio, I.; Liu, J.; Janssen, J.M.; Chen, X.; Gonçalves, M.A. Adenoviral vectors encoding CRISPR/Cas9 multiplexes rescue dystrophin synthesis in unselected populations of DMD muscle cells. *Sci. Rep.* **2016**, *6*, 37051.
111. Brescia, M.; Janssen, J.M.; Liu, J.; Gonçalves, M.A. High-capacity adenoviral vectors permit robust and versatile testing of DMD gene repair tools and strategies in human cells. *Cells* **2020**, *9*, 869.
112. Li, C.; Guan, X.; Du, T.; Jin, W.; Wu, B.; Liu, Y.; Wang, P.; Hu, B.; Griffin, G.E.; Shattock, R.J.; et al. Inhibition of HIV-1 infection of primary CD4+ T-cells by gene editing of CCR5 using adenovirus-delivered CRISPR/Cas9. *J. Gen. Virol.* **2015**, *96*, 2381–2393.

113. Li, C.; Psatha, N.; Sova, P.; Gil, S.; Wang, H.; Kim, J.; Kulkarni, C.; Valensisi, C.; Hawkins, R.D.; Stamatoyannopoulos, G.; et al. Reactivation of gamma-globin in adult beta-YAC mice after ex vivo and in vivo hematopoietic stem cell genome editing. *Blood* **2018**, *131*, 2915–2928.
114. Li, C.; Psatha, N.; Gil, S.; Wang, H.; Papayannopoulou, T.; Lieber, A. HDAd5/35(++) Adenovirus Vector Expressing Anti-CRISPR Peptides Decreases CRISPR/Cas9 Toxicity in Human Hematopoietic Stem Cells. *Mol. Ther. Methods Clin. Dev.* **2018**, *9*, 390–401.
115. Schiroli, G.; Conti, A.; Ferrari, S.; Della Volpe, L.; Jacob, A.; Albano, L.; Beretta, S.; Calabria, A.; Vavassori, V.; Gasparini, P.; et al. Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* **2019**, *24*, 551–565.
116. Coluccio, A.; Miselli, F.; Lombardo, A.; Marconi, A.; Malagoli Tagliazucchi, G.; Gonçalves, M.A.; Pincelli, C.; Maruggi, G.; Del Rio, M.; Naldini, L.; et al. Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther.* **2013**, *21*, 1695–1704.
117. 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.
118. Li, C.; Mishra, A.S.; Gil, S.; Wang, M.; Georgakopoulou, A.; Papayannopoulou, T.; Hawkins, R.D.; Lieber, A. Targeted Integration and High-Level Transgene Expression in AAVS1 Transgenic Mice after In Vivo HSC Transduction with HDAd5/35++ Vectors. *Mol. Ther.* **2019**, *27*, 2195–2212.
119. Thomson, J.A.; Itskovitz-Eldor, J.; Shapiro, S.S.; Waknitz, M.A.; Swiergiel, J.J.; Marshall, V.S.; Jones, J.M. Embryonic stem cell lines derived from human blastocysts. *Science* **1998**, *282*, 1145–1147.
120. de Miguel-Beriain, I. The ethics of stem cells revisited. *Adv. Drug Deliv. Rev.* **2015**, *82-83*, 176-180.
121. Hu, J.; Wang, J. From embryonic stem cells to induced pluripotent stem cells-Ready for clinical therapy? *Clin. Transplant* **2019**, *33*, e13573.
122. Takahashi, K.; Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **2006**, *126*, 663–676.
123. Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K.; Yamanaka, S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **2007**, *131*, 861–872.
124. Doss MX, Sachinidis, A. Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. *Cells* **2019**, *8*, 403.
125. Kimbrel, E.A.; Lanza, R. Pluripotent stem cells: The last 10 years. *Regen. Med.* **2016**, *11*, 831–847.
126. Hockemeyer, D.; Jaenisch, R. Induced Pluripotent Stem Cells Meet Genome Editing. *Cell Stem Cell* **2016**, *18*, 573–586.
127. Li, M.; Suzuki, K.; Kim, N.Y.; Liu, G.H.; Izpisua Belmonte, J.C. A cut above the rest: Targeted genome editing technologies in human pluripotent stem cells. *J. Biol. Chem.* **2014**, *289*, 4594–4599.
128. Chen, X.; Janssen, J.M.; Liu, J.; Maggio, I.; t Jong, A.E.J.; Mikkers, H.M.M.; Gonçalves, M. In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun*. **2017**, *8*, 657.
129. Nakajima, K.; Zhou, Y.; Tomita, A.; Hirade, Y.; Gurumurthy, C.B.; Nakada, S. Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res*. **2018**, *28*, 223–230.
130. Hyodo, T.; Rahman, M.L.; Karnan, S.; Ito, T.; Toyoda, A.; Ota, A.; Wahiduzzaman, M.; Tsuzuki, S.; Okada, Y.; Hosokawa, Y.; et al. Tandem Paired Nicking Promotes Precise Genome Editing with Scarce Interference by p53. *Cell Rep*. **2020**, *30*, 1195–1207.
131. Rees, H.A.; Liu, D.R. Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **2018**, *19*, 770–788.
132. Anzalone, A.V.; Randolph, P.B.; Davis, J.R.; Sousa, A.A.; Koblan, L.W.; Levy, J.M.; Chen, P.J.; Wilson, C.; Newby, G.A.; Raguram, A.; et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019**, *576*, 149–157.

133. Haapaniemi, E.; Botla, S.; Persson, J.; Schmierer, B.; Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **2018**, *24*, 927–930.
134. Ihry, R.J.; Worringer, K.A.; Salick, M.R.; Frias, E.; Ho, D.; Theriault, K.; Kommineni, S.; Chen, J.; Sondey, M.; Ye, C.; et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **2018**, *24*, 939–946.
135. Gonçalves, M.A.; van Nierop, G.P.; 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.
136. Rui, Y.; Wilson, D.R.; Green, J.J. Non-Viral Delivery To Enable Genome Editing. *Trends Biotechnol*. **2019**, *37*, 281–293.
137. Ohbayashi, F.; Balamotis, M.A.; Kishimoto, A.; Aizawa, E.; Diaz, A.; Hasty, P.; Graham, F.L.; Caskey, C.T.; Mitani, K. Correction of chromosomal mutation and random integration in embryonic stem cells with helper-dependent adenoviral vectors. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13628–13633.
138. Suzuki, K.; Mitsui, K.; Aizawa, E.; Hasegawa, K.; Kawase, E.; Yamagishi, T.; Shimizu, Y.; Suemori, H.; Nakatsuji, N.; Mitani, K. Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13781–13786.
139. Aizawa, E.; Hirabayashi, Y.; Iwanaga, Y.; Suzuki, K.; Sakurai, K.; Shimoji, M.; Aiba, K.; Wada, T.; Tooi, N.; Kawase, E.; et al. Efficient and accurate homologous recombination in hESCs and hiPSCs using helper-dependent adenoviral vectors. *Mol. Ther*. **2012**, *20*, 424–431.
140. Umeda, K.; Suzuki, K.; Yamazoe, T.; Shiraki, N.; Higuchi, Y.; Tokieda, K.; Kume, K.; Mitani, K.; Kume, S. Albumin gene targeting in human embryonic stem cells and induced pluripotent stem cells with helper-dependent adenoviral vector to monitor hepatic differentiation. *Stem Cell Res.* **2013**, *10*, 179–194.
141. Sone, T.; Shin, M.; Ouchi, T.; Sasanuma, H.; Miyamoto, A.; Ohte, S.; Tsukamoto, S.; Nakanishi, M.; Okano, H.; Katagiri, T.; et al. Dual usage of a stage-specific fluorescent reporter system based on a helper-dependent adenoviral vector to visualize osteogenic differentiation. *Sci. Rep.* **2019**, *9*, 9705.
142. Liu, G.H.; Suzuki, K.; Qu, J.; Sancho-Martinez, I.; Yi, F.; Li, M.; Kumar, S.; Nivet, E.; Kim, J.; Soligalla, R.D.; et al. Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. *Cell Stem Cell* **2011**, *8*, 688–694.
143. Li, M.; Suzuki, K.; Qu, J.; Saini, P.; Dubova, I.; Yi, F.; Lee, J.; Sancho-Martinez, I.; Liu, G.H.; Izpisua Belmonte, J.C. Efficient correction of hemoglobinopathy-causing mutations by homologous recombination in integration-free patient iPSCs. *Cell Res.* **2011**, *21*, 1740–1744.
144. Palmer, D.J.; Grove, N.C.; Ing, J.; Crane, A.M.; Venken, K.; Davis, B.R.; Ng, P. Homology Requirements for Efficient, Footprintless Gene Editing at the CFTR Locus in Human iPSCs with Helper-dependent Adenoviral Vectors. *Mol. Ther. Nucleic Acids* **2016**, *5*, e372.
145. Palmer, D.J.; Grove, N.C.; Turner, D.L.; Ng, P. Gene Editing with Helper-Dependent Adenovirus Can Efficiently Introduce Multiple Changes Simultaneously over a Large Genomic Region. *Mol. Ther. Nucleic Acids* **2017**, *8*, 101–110.
146. Liu, G.H.; Qu, J.; Suzuki, K.; Nivet, E.; Li, M.; Montserrat, N.; Yi, F.; Xu, X.; Ruiz, S.; Zhang, W.; et al. Progressive degeneration of human neural stem cells caused by pathogenic LRRK2. *Nature* **2012**, *491*, 603–607.
147. Liu, G.H.; Suzuki, K.; Li, M.; Qu, J.; Montserrat, N.; Tarantino, C.; Gu, Y.; Yi, F.; Xu, X.; Zhang, W.; et al. Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. *Nat. Commun.* **2014**, *5*, 4330.
148. Yoshida, T.; Ozawa,,Y.; Suzuki, K.; Yuki, K.; Ohyama, M.; Akamatsu, W.; Matsuzaki, Y.; Shimmura, S.; Mitani, K.; Tsubota, K.; et al. The use of induced pluripotent stem cells to reveal pathogenic gene mutations and explore treatments for retinitis pigmentosa. *Mol. Brain* **2014**, *7*, 45.
149. Zhang, W.; Li, J.; Suzuki, K.; Qu, J.; Wang, P.; Zhou, J.; Liu, X.; Ren, R.; Xu, X.; Ocampo, A.; et al. Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* **2015**, *348*, 1160–1163.
150. Suzuki, K.; Yu, C.; Qu, J.; Li, M.; Yao, X.; Yuan, T.; Goebl, A.; Tang, S.; Ren, R.; Aizawa, E.; et al. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell* **2014**, *15*, 31–36.

151. Miller, D.G.; Petek, L.M.; Russell, D.W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **2004**, *36*, 767.
152. Hanlon, K.S.; Kleinstiver, B.P.; Garcia, S.P.; Zaborowski, M.P.; Volak, A.; Spirig, S.E.; Muller, A.; Sousa, A.A.; Tsai, S.Q.; Bengtsson, N.E. High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat. Commun.* **2019**, *10*, 1–11.
153. Komor, A.C.; Kim, Y.B.; Packer, M.S.; Zuris, J.A.; Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016**, *533*, 420–424.
154. Gaudelli, N.M.; Komor, A.C., Rees, H.A.; Packer, M.S., Badran; A.H., Bryson, D.I.; Liu, D.R. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **2017**, *551*, 464–471.
155. Kosicki, M.; Tomberg, K.; Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol*. **2018**, *36*, 765–771.
156. Park, S.; Beal, P.A. Off-Target Editing by CRISPR-Guided DNA Base Editors. *Biochemistry* **2019**, *58*, 3727–3734.

Chapter 2

Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on *in trans* paired nicking

Xiaoyu Chen[1,†], Francesca Tasca[1,†], Qian Wang[1], Jin Liu[1], Josephine M. Janssen[1], Marcella D. Brescia[1], Milena Bellin[2], Karoly Szuhai[1], Josefin Kenrick[3], Richard L. Frock[3] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands;
[2]Department of Anatomy and Embryology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands;
[3]Department of Radiation Oncology, Division of Radiation and Cancer Biology, Stanford University School of Medicine, 269 Campus Dr. Stanford, CA 94305, USA
[†]These authors contributed equally

Nucleic Acids Research, 2020, Vol.48, No.2, ***974–995***
doi:10.1093/nar/gkz1121

ABSTRACT

Genome editing typically involves recombination between donor nucleic acids and acceptor genomic sequences subjected to double-stranded DNA breaks (DSBs) made by programmable nucleases (e.g. CRISPR–Cas9). Yet, nucleases yield off-target mutations and, most pervasively, unpredictable target allele disruptions. Remarkably, to date, the untoward phenotypic consequences of disrupting allelic and non-allelic (e.g. pseudogene) sequences have received scant scrutiny and, crucially, remain to be addressed. Here, we demonstrate that gene-edited cells can lose fitness as a result of DSBs at allelic and non-allelic target sites and report that simultaneous single-stranded DNA break formation at donor and acceptor DNA by CRISPR–Cas9 nickases (*in trans* paired nicking) mostly overcomes such disruptive genotype-phenotype associations. Moreover, *in trans* paired nicking gene editing can efficiently and precisely add large DNA segments into essential and multiple-copy genomic sites. As shown herein by genotyping assays and high-throughput genome-wide sequencing of DNA translocations, this is achieved while circumventing most allelic and non-allelic mutations and chromosomal rearrangements characteristic of nuclease-dependent procedures. Our work demonstrates that *in trans* paired nicking retains target protein dosages in gene-edited cell populations and expands gene editing to chromosomal tracts previously not possible to modify seamlessly due to their recurrence in the genome or essentiality for cell function.

INTRODUCTION

Genome editing based on homology-dependent and homology-independent DNA repair pathways activated by programmable nucleases permits modifying specific chromosomal sequences in living cells (1). Importantly, these genetic changes can span from single base pairs to whole transgenes (2). However, the genomic double-stranded DNA breaks (DSBs) required for DNA repair activation inevitably yield complex and unpredictable genetic structural variants. These by-products result from the fact that DSBs (targeted or otherwise) are substrates for prevalent non-homologous end joining (NHEJ) pathways and other error-prone recombination processes (3). These processes can trigger local (4) and genome-wide mutations and rearrangements, in the form of insertions and deletions (indels), duplications and/or translocations (5–10). Likewise insidious, targeted DSBs at homologous alleles can result in the assembly of unstable dicentric chromosomes through head-to-head inversional translocations (10). Finally, the engagement of donor DNA with target and off-target DSBs often leads to inaccurate and random chromosomal insertion events, respectively (2,11). This is especially so when donor DNA is presented in target cell nuclei as free-ended double-stranded recombination substrates (11–13).

The unpredictability of genome editing outcomes is naturally aggravated whenever nuclease target sites are located in (i) coding sequences, especially those associated with essentiality and haploinsufficiency, (ii) overlapping *trans*-acting or *cis*-acting sequences and (iii) multiple-copy sequences, such as those in paralogs and pseudogenes. To date, genotypic and phenotypic consequences resulting from editing these three types of genomic regions have received limited examination and remain to be addressed.

Single-stranded DNA breaks (SSBs) made by programmable sequence-specific and strand-specific nucleases (nickases) are intrinsically less disruptive than DSBs as they do not constitute canonical NHEJ substrates (14–17). In this regard, CRISPR–Cas9 nickases consisting of guide RNAs (gRNAs) and Cas9 proteins with either their RuvC or HNH nuclease domains disabled (e.g. $Cas9^{D10A}$ and $Cas9^{H840A}$, respectively), are particularly appealing programmable nicking enzymes (18–20). Indeed, similarly to their cleaving counterparts, CRISPR-Cas9 nickases target DNA consisting of a protospacer adjacent motif (PAM; NGG in *Streptococcus pyogenes* SpCas9) and a sequence complementary to the 5'-terminal 20 nucleotides (nts) of the gRNA (spacer) (18,21). Pairs of CRISPR–Cas9 nickases are commonly used to induce site-specific DSBs through coordinated nicking at opposite target DNA strands. This dual nicking strategy can significantly improve the specificity of DSB formation as SSBs made at off-target sites are, for the most part, faithfully repaired (22,23). However, genome editing based on paired CRISPR–Cas9 nickases remains prone to mutagenesis and chromosomal rearrangements due to the ultimate creation of DSBs (12,22,23).

The non-disruptive character of genome editing based on targeted chromosomal SSBs offers the possibility for seamlessly modifying a broad range of genomic sequences, including those that encode functional protein motifs or essential proteins or that are present in genomic tracts with high similarity

to DNA located elsewhere in the genome. Unfortunately, chromosomal SSBs are, *per se*, poor stimuli for genome editing via precise homology-directed DNA repair (HDR), even in instances in which single base pairs are due to be inserted at a target site (14–17,24).

Here, we sought to determine whether chromosomal regions previously not possible to edit in an efficient and seamless manner could in fact be modified as such. In particular, we hypothesized that *in trans* paired nicking, comprising coordinated SSB formation at donor and acceptor HDR substrates by CRISPR–Cas9 nickases, permits expanding the 'editable genome', i.e. the genomic space amenable to operative DNA editing. Recently, it has been demonstrated that this genetic engineering principle achieves precise HDR-mediated genomic insertions, from a few base pairs (12,25) to whole transgenes (12), without provoking the competing NHEJ pathway. However, the performance of *in trans* paired nicking at coding sequences of endogenous genes, in particular those associated with haploinsufficiency and essentiality, is unknown. To date, equally unknown is the performance of genome editing approaches based on repairing SSBs versus DSBs at these coding sequences using donor plasmids. By targeting exons in the *H2A.X variant histone gene* (*H2AX*) and the *POU class 5 homebox 1 gene* (*POU5F1* or *OCT4*), whose products are essential for the DNA damage response and stem cell pluripotency, respectively, we demonstrate that in contrast to DSB-dependent strategies, in trans paired nicking achieves precise gene editing while disrupting neither functional motifs nor allelic or non-allelic homologous DNA. Moreover, after adapting linear amplification-mediated high-throughput genome-wide translocation sequencing (HTGTS) (10,26) for the detection of SSB-initiated translocations, we found that CRISPR-SpCas9 nickases greatly reduce large-scale chromosomal rearrangements when compared to their nuclease counterparts. Finally, *PARP1* gene targeting experiments showed that, also in instances in which a target gene is not associated with haploinsufficiency or essentiality, in trans paired nicking achieves accurate HDR-mediated gene knock-ins without mutagenizing unmodified alleles, and hence, without reducing target protein dosages.

MATERIALS AND METHODS

Cells

Human cervix carcinoma HeLa cells and human embryonic kidney 293T (HEK293T) cells (both from American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (DMEM; ThermoFisher Scientific; Cat. No.: 41966029) supplemented with 5% (v/v) and 10% (v/v), respectively, fetal bovine serum ultra-low endotoxin (FBS; biowest; Cat. No.: S1860500). The HeLa cells, authenticated before by karyotyping analysis (11), were used for gene editing experiments. The HEK293T cells were used for assembling lentiviral vector LV.Cre particles and orthogonal HTGTS analyses. The generation and characterization of the human induced pluripotent stem cells (iPSCs) used in this work (LUMC0020iCTRL) were detailed elsewhere (27). In the current study, these cells were further characterized by COBRA-FISH karyotyping. The iPSCs were cultured in feeder-free Essential 8 Medium (E8; ThermoFisher Scientific; Cat. No.: A1517001) supplemented with 25 U ml^{-1} penicillin and 25 µg ml^{-1} of streptomycin (ThermoFisher Scientific; Cat. No.: 15140122). The iPSCs were kept in wells of six-well plates (Greiner Bio-One; Cat. No.: 662160) coated for 1 h at room with Vitronectin Recombinant Human Protein (VTN-N; ThermoFisher Scientific; Cat. No.: A14700) diluted 1:100 to a final concentration of 5 ng ml^{-1} in Dulbecco's phosphate-buffered saline, no calcium, no magnesium (DPBS; ThermoFisher Scientific; Cat. No.: 14190094). When ready for sub-culturing, to let cell-cell dissociation occur, the iPSCs were first washed with DPBS solution and then incubated with 0.5 mM ethylenediaminetetraacetic acid (EDTA; Invitrogen Cat. No.: 15575020) in DPBS at 37°C and room temperature for 4 and 1 min, respectively. After the removal of the EDTA solution, the cells were seeded in new wells of 24-well plates coated with VTN-N and containing E8 medium supplemented with a 1:200 dilution of RevitaCell (ThermoFisher Scientific; Cat. No.: A2644501). The cells used in this study were mycoplasma free and were kept at 37°C in a humidified-air atmosphere with 5% CO_2 (iPSCs) or 10% CO_2 (HeLa and HEK293T cells).

Recombinant DNA

The expression plasmids AU26_pCAG.Cas9 and AU28_pCAG.Cas9^{D10A} encoding cleaving Cas9 and nicking Cas9^{D10A} enzymes, respectively, have been described previously (12). The control plasmid gRNA_Cloning Vector (Addgene #41824) and the *OCT4*-targeting donor construct eGFP-PGK-Puro (Addgene #31937), herein named pgRNA^{Empty} and pDonor^{OCT4}, respectively, were also described before (20,28). The annotated maps and nucleotide sequences of donor constructs AX74_pDonor$^{OCT4.TS}$,

AX66_pDonor$^{OCT4.1TS}$, AZ44_pDonor^{H2AX}, AZ25_pDonor$^{H2AX.TS}$, AW77_pDonor^{PARP1} and AW69_pDonor$^{PARP1.TS}$ are available in pages 1-14 of the **Supplementary Information**. The annotated maps
and nucleotide sequences of the *S. pyogenes* gRNA-expressing plasmids AZ34_pgRNA$^{H2AX.1}$, AZ35_pgRNA$^{H2AX.2}$, AM70_pgRNA^{PARP1}, AX33_pgRNA$^{OCT4.1}$, AX34_pgRNA$^{OCT4.2}$ are available in pages 15-24 of the **Supplementary Information.** The annotated map and nucleotide sequence of the Cre-expressing lentiviral vector construct BC17_pLV.Cre is available in pages 25-27 of the **Supplementary Information**. The constructs used in the experiments for identifying CRISPR-SaCas9 nucleases inducing HTGTS bait DSBs at RAG1 were BA15_pCAG.SaCas9.rBGpA (29), AV85_pSa-gRAG1.1, AV86_pSa-gRAG1.2, AV87_pSa-gRAG1.3, AP65_pSa-gAAVS1. With the exception of BA15_pCAG.SaCas9.rBGpA (29), all these constructs are described in pages 28–33 of the **Supplementary Information**. The plasmid BPK2660 (Addgene #70709) served as a negative control as it encodes an irrelevant, non-targeting, Staphylococcus aureus gRNA, herein named Sa-gNT (30). Moreover, after BsmBI digestion, BPK2660 also served as an isogenic cloning vector for the insertion of annealed oligonucleotides corresponding to the spacers of S. aureus gRNAs; Sa-gRAG1.1, Sa-gRAG1.2, Sa-gRAG1.3 and Sa-gAAVS1. Plasmids encoding *S. aureus* CRISPR components used for inducing universal HTGTS bait DSBs (i.e., BA15_pCAG.SaCas9.rBGpA and AV85_pSa-gRAG1.1), were combined with constructs AV62_pCAG.Cas9.rBGpA, AB65_pCAG.Cas9^{D10A}.rBGpA and gRNA_AAVS1-T2 (20) expressing *S. pyogenes* CRISPR elements for triggering test HTGTS prey DNA lesions in the form of *AAVS1*-targeted DSBs or SSBs. The latter plasmid (Addgene #41818) encodes an *AAVS1*-targeted gRNA, herein dubbed gAAVS1. The annotated maps and nucleotide sequences of AV62_pCAG.Cas9.rBGpA and AB65_pCAG.Cas9^{D10A}.rBGpA are described in pages 34-39 of the **Supplementary Information**. The full sequences and annotated maps of the plasmids applied in the *AAVS1* gene targeting experiments; AV15_pCAG.Cas9.gRNA^{S1}, AV44_pCAG.Cas9^{D10A}.gRNA^{S1}, AV13_pCAG.Cas9.gRNA^{NT}, AV11_pDonor.EP^{S1} (Addgene #100296) and AV09_pDonor.EP$^{S1.TS}$ (Addgene #100297) are available elsewhere (12).

HeLa and HEK293T cell transfections

HeLa and HEK293T cells were seeded in the tissue culture vessels indicated in **Supplementary Tables S1–S6**. The next day, transfections started by adding a 1 mg ml^{-1} 25 kDa linear polyethyleneimine (PEI, Polysciences) solution (pH 7.4) to each plasmid mixture diluted in 50 μl of 150 mM NaCl (Merck). The cell numbers, the amounts of PEI and DNA (in ng) as well as the compositions of each of the DNA mixtures corresponding to the different transfection reactions are specified in **Supplementary Tables S1–S6**. After the addition of PEI, the transfection reactions were immediately and vigorously vortexed for 10 s, after which, DNA-PEI complexes were allowed to form for 15 min at room temperature. The resulting DNA-PEI complexes were subsequently added directly into the culture media of the target cells and, after 6 h, the transfection media were substituted by regular culture media. Whenever appropriate, reporter-directed flow cytometry was performed at 3 days post-transfection to determine the transfection efficiencies. In the gene targeting experiments, cell populations were then sub-cultured for at least 2 weeks to eliminate episomal donor DNA templates, after which, reporter-directed flow cytometry was used to quantify the frequencies of stably transfected cells.

Human induced pluripotent stem cell (iPSCs) transfections

The iPSCs were first seeded in wells of 24-well plates (Greiner Bio-One) that had been previously coated with VTN-N (ThermoFisher Scientific) as indicated above. The next day, the iPSC culture media were refreshed at least 2 h prior to transfection. Transfections were initiated by adding the appropriate plasmid mixtures together with Lipofectamine Stem Transfection Reagent (ThermoFisher Scientific, Cat. No.: STEM00003) to 50 μl of Opti-MEM medium (Gibco; Cat. No.: 31985-047) in 1.5-ml sterile Eppendorf tubes (**Supplementary Tables S7** and **S8**). After mixing by pipetting, the transfection reactions were incubated at room temperature for 10 min and were then added into the culture media of the target iPSCs (**Supplementary Tables S7** and **S8**). The media were replaced 24 h later and, at 2–3 days post-transfection, the iPSCs were transferred into a new culture well and were subsequently expanded in wells of 6-well plates (Greiner Bio-One) for 5–7 days in the presence of 0.5 μg ml−1 puromycin in StemFlex Medium (ThermoFisher Scientific, Cat. No.: A3349401) containing 25 U ml^{-1} penicillin and 25 μg ml^{-1} of streptomycin. Parallel cultures of mock-transfected iPSCs served as negative controls. At the end of the selection period, puromycin-resistant iPSC colonies were identified by using the leukocyte alkaline phosphatase kit (Sigma-Aldrich; Cat. No.: 86R-1KT) for detecting enzymatic activity from the

pluripotency marker alkaline phosphatase. Cultures of puromycin-resistant iPSC populations and individual randomly selected iPSC colonies were also expanded, collected and cryopreserved for further analyses. The iPSC genomic DNA samples used for orthogonal HTGTS analyses were generated by nucleofecting iPSCs with constructs expressing SaCas9:Sa-gRAG1.1 and SpCas9:gAAVS1 or SaCas9:Sa-gRAG1.1 and SpCas9^{D10A}:gAAVS1. Nucleofection of iPSCs with plasmids expressing only the SaCas9:Sa-gRAG1.1 complexes needed for generating bait DSBs served as an orthogonal HTGTS assay control (**Supplementary Table S9**). The iPSC nucleofections were performed in a Nucleofector 2b-device (Lonza) using Amaxa Human Stem Cell Nucleofector Kit 2 (Lonza; Cat. No.: VPH-5022). A total amount of 8 μg of DNA diluted in 10 μl of Milli-Q water were added to 100 μl of nucleofection buffer containing 2 × 10^6 iPSCs. After gentle mixing, the cell suspensions were transferred to the device-tailored cuvettes and immediately subjected to the nucleofection program B-016, selected for human embryonic stem cells. Next, the iPSCs were transferred to wells of 6-well plates (Greiner Bio-One) containing 2 ml of pre-warmed E8 medium (ThermoFisher Scientific; Cat. No.: A1517001) supplemented with a 1:100 dilution of RevitaCell (ThermoFisher Scientific; Cat. No.: A2644501). After an overnight incubation period, the culture medium was replenished and, at 3 days post-nucleofection, genomic DNA was extracted. Finally, genomic DNA samples were subjected to T7 endonuclease I (T7EI)-based genotyping assays directed at *RAG1* and *AAVS1* alleles and, subsequently, orthogonal HTGTS analyses was performed as described below.

Orthogonal HTGTS sample preparation

Transfections for generating genomic DNA samples for orthogonal HTGTS analyses were carried out in HEK293T cells and iPSCs (**Supplementary Tables S1** and **S9**, respectively). The genomic DNA was isolated at 36 h post-transfection as described before (31). In brief, the cells were collected by centrifugation and resuspended in lysis buffer consisting of 200 mM NaCl, 10 mM Tris–HCl pH 7.4, 2 mM EDTA pH 8.0, 0.2% (w/v), sodium dodecyl sulphate (SDS) and freshly added proteinase K (Thermo Fisher Scientific; Cat. No.: #EO0491) at a final concentration of 200 ng ml^{-1}. After an overnight incubation period at 56°C, the DNA was precipitated by adding isopropanol (1:1) and immediate mixing of the aqueous and organic phases. Next, the DNA was transferred to a new Eppendorf tube containing 1 ml of 70% (v/v) ethanol. The DNA was next pelleted by centrifugation at 13 000 × g for 5 min at 4°C, and dissolved in TE buffer (10 mM Tris–HCl pH 8.0; 1 mM EDTA pH 8.0) for at least 2 h at 56°C. Before orthogonal HTGTS analyses, genomic DNA samples were subjected to T7EI-based genotyping assays. These assays permitted assessing bait and prey chromosomal DNA breaks at *RAG1* and *AAVS1* alleles, respectively, in HEK293T and iPSC cell populations. To this end, the *RAG1* and *AAVS1* target regions were PCR-amplified with KOD Hot Start DNA Polymerase (Merck Millipore; Cat. No.: 71086–3) and GoTaq G2 Flexi DNA Polymerase (Promega; Cat. No.: M7805) using the PCR mixtures indicated in **Supplementary Tables S10** and **S11**, respectively. The PCR primers and cycling parameters used to amplify *RAG1* and *AAVS1* DNA are specified in **Supplementary Tables S12** and **S13**, respectively. Indels generated by NHEJ-mediated DSB repair were detected by exposing *RAG1* and *AAVS1* amplicons to T7EI (Biolabs; Cat. No.: M0302L) as below indicated. Transfections for selecting Sa-gRNAs inducing universal HTGTS bait DSBs at *RAG1* were performed on HeLa cells and HEK293T cells (**Supplementary Table S2**). At 3 days post-transfection, indel formation at the target gene was assessed by T7EI-based genotyping assays as below indicated. To this end, genomic DNA was extracted by using the DNeasy Blood & Tissue kit (Qiagen; Cat. No.: 69506) according to the manufacturer's instructions. Next, the RAG1 target region in HeLa and HEK293T cells was PCR-amplified with KOD Hot Start DNA Polymerase (Merck Millipore). The PCR mixtures, primers and cycling parameters are indicated in **Supplementary Tables S10**, **S12** and **S13**, respectively. The construct expressing *S. aureus* gRNA Sa-gRAG1.1 was selected to induce bait DSBs at *RAG1* in orthogonal HTGTS experiments in HEK293T cells and iPSCs (**Supplementary Tables S1** and **S9**, respectively).

Gene targeting and gene tagging experiments

Transfections for *AAVS1* gene targeting experiments were done in HEK293T cells and iPSCs (**Supplementary Tables S3** and **S8**, respectively) using as donors plasmids AV11_pDonor.EP^{S1} (Addgene #100296) and AV09_pDonor.EP$^{S1.TS}$ (Addgene #100297) (12). The former differs from the latter in that is has its targeting module flanked by gAAVS1 target sites. The targeting modules of these donors consist of sequences homologous to the AAVS1 locus framing expression units encoding both puromycin N-acetyltransferase and EGFP. In these experiments, these donors were combined with plasmids AV15_pCAG.Cas9.gRNA^{S1}, AV44_pCAG.Cas9^{D10A}.gRNA^{S1} and AV13_pCAG.Cas9.gRNA^{NT}

which co-express SpCas9 proteins and gRNAs (12). At 3 days post-transfection, the transfection efficiencies were determined by EGFP-directed flow cytometry. Subsequently, the cells were sub-cultured for 14 days, for the removal of episomal donor templates, after which stable transfection frequencies were established via EGFP-directed flow cytometry. In addition, stably transfected cells present in long-term HEK293T cell cultures were selected for by incubation with 3 μg ml^{-1} of puromycin (InvivoGen; Cat. No.: 58582) during 9 days. The distribution of EGFP expression levels in the resulting puromycin-resistant populations was assessed by EGFP-directed flow cytometry. Transfections for tagging H2AX and PARP1 proteins were performed on HeLa cells (**Supplementary Tables S4** and **S5**, respectively). Transfections of HeLa cells for *OCT4* gene targeting (**Supplementary Table S6**), were assessed by colony-formation assays. To this end, at approximately 2 weeks post-transfection, the cells were counted and seeded at a density of 10^5 cells per 60 mm × 15 mm culture dishes (Greiner Bio-One; Cat. No.: 628160). After a 17-day exposure period to 1 μg ml^{-1} of puromycin (InvivoGen), HeLa cell colonies were identified by Giemsa staining.

Determining genome-wide off-target effects by orthogonal HTGTS analysis

The orthogonal HTGTS analyses were done in a blind fashion on genomic DNA samples isolated from HEK293T cells and iPSCs. Genomic DNA samples from the former and latter cell types were generated as described above using the transfection mixtures specified in **Supplementary Tables S1** and **S9**, respectively. The reagents and procedures for HTGTS analysis have been detailed elsewhere (10,31). In brief, 25 μg of genomic DNA was used for each sample. Samples were sheared using a Bioruptor (Diagenode) with a circulating temperature of 4°C, on a low power setting: 2 × 30 s pulses interspaced by a 60 s cool down period. The biotinylated RAG1A/B – F1 primer (10) was used for LAM-PCR (31), and ssDNA products were enriched on streptavidin-coated magnetic beads (ThermoFisher Scientific; Cat. No.: 65002) prior to ligation of bridge adapters (10,31). Barcoded RAG1A/B – F2 I5 and AP2 I7 primers (10) were used for the nested PCR. P5–I5 and P7–I7 primers (31) were used in the final PCR. The resulting amplicons between 500 bp to 1 kb were separated and gel extracted (Qiagen; Cat. No.: 28706). Phusion polymerase (ThermoFisher Scientific; Cat. No.: F530L) was used in all PCR steps and the blocking enzyme step was omitted. HTGTS libraries were run on a Bioanalyzer (Agilent 2100) prior to MiSeq 2 × 250 bp sequencing (Illumina; Cat. No.: MS-102-2003). Pooled sequence reads were demultiplexed and trimmed according to predetermined molecular barcodes and adapter sequences; each library was subjected to bait/prey alignments (hg19), filtering, and post-pipeline analysis as described (31). Significantly enriched translocation sites in sequence read libraries from individual experiments were identified using MACS2 as previously described (10). Translocation hotspots were called if such enriched translocation sites were statistically significant in the majority of the independent replicate experiments.

Characterization of genome-modifying events by clonal analysis

$EGFP^+$ and $mCherry^+$ HeLa cells generated after PARP1 and H2AX gene editing, respectively, were sorted at 2–3 weeks post-transfection as single cells or as whole populations with the aid of a BD FACSAria III flow cytometer (BD Biosciences). The single cell-derived clones were seeded in wells of 96-well plates (Greiner Bio-One) and were grown in HeLa culture medium supplemented with 50 U ml^{-1} penicillin, 50 μg ml^{-1} of streptomycin and, to increase their cloning efficiency, 50 μM α-thioglycerol and 20 nM bathocuproine disulfonate (both from Sigma-Aldrich) (32). Next, conventional and junction PCR analyses were performed on chromosomal DNA from individual clones, each of which representing a specific genome-modifying event. The PCR screening of the $mCherry^+$ HeLa cell clones was done with the GoTaq G2 Flexi DNA Polymerase system (Promega; Cat. No.: M7808) using the PCR mixtures and cycling parameters indicated in **Supplementary Tables S14** and **S15**, respectively. The screening of the $EGFP^+$ HeLa cell clones was performed with the reagents and protocol provided in the Phire Tissue Direct PCR Master Mix kit (ThermoFisher Scientific, Cat. No.: F170L). The PCR mixtures and cycling parameters used for these analyses are also indicated in the **Supplementary Tables S14** and **S15**, respectively.

Characterization of genome-modifying events in iPSCs by clonal analysis

Puromycin-resistant iPSC colonies derived from *OCT4* targeting experiments using $pDonor^{OCT4}$ and $pDonor^{OCT4.TS}$, were picked from 6-well plates and transferred into wells of 96-well plates by applying a standard 'cut-and-paste' technique. The resulting iPSC clones, each of which representing an individual genome-modifying event, were first cultured in StemFlex Medium (ThermoFisher Scientific) containing

25 U ml^{-1} penicillin and 25 μg ml^{-1} of streptomycin supplemented with Revitacell (ThermoFisher Scientific). Next, the iPSC clones were expanded and adapted to E8 medium (ThermoFisher Scientific) in wells of 24-well plates (Greiner-BioOne). The junction PCR screening for detecting and characterizing genome-modifying events in iPSCs was done on total genomic DNA purified from iPSC clones using the reagents and protocol provided in the Phire Tissue Direct PCR Master Mix kit (ThermoFisher Scientific). The PCR mixtures and cycling parameters applied for these analyses are indicated in the **Supplementary Tables S14** and **S15**, respectively.

Characterization of iPSC clones by COBRA-FISH analysis

Combined binary ratio labelling (COBRA) multicolour FISH-based molecular karyotyping (COBRA-FISH) was carried out on native and gene-edited iPSC lines essentially as detailed elsewhere (33). In brief, glass coverslips containing metaphase spreads air-dried for at least 24 h were incubated with 100 μg ml^{-1} RNase A (Roche; Cat. No.: 10154105103) in 2× saline-sodium citrate (SSC; Sigma-Aldrich; Cat. No.: S0902) at 37 °C for 10 min, followed by incubation with 0.005% pepsin (Sigma-Aldrich; Cat. No.: P0525000) in 0.1 M HCl for 5 min at 37 °C and fixation with 1% formaldehyde (Merck; Cat. No.: 1.03999.1000) in PBS pH 7.4 at room temperature for 10 min. The specimens were dehydrated through a series of incubations in 70–90–100% ethanol solutions, 3 min each, followed by air drying. The probe mix containing the paint mixes covering all chromosomes was dissolved in hybridization mixture, denatured and let hybridize in a moist chamber for 72 h. After hybridization, the glass coverslips were washed in 2× SSC and 0.1% Tween-20 (Promega, Cat. No.: PRH5152), then in 50% formamide (Merck; Cat. No.: 1.09684.1000), 2× SSC pH 7.0 solution at 44°C followed by incubation in 0.1× SSC at 60°C. Each washing step was performed twice for 5 min. The specimens were then dehydrated through a series of incubations in 70–90–100% ethanol solutions, air-dried and embedded in Citifluor AF1/DAPI (400 ng ml^{-1}) solution (Aurion; Cat. No.: E17970). Stained chromosomes were visualised using a Leica DMRA fluorescence microscope (Leica, Wetzlar, Germany) and images were captured with the aid of a CoolSnap HQ2 camera (Photometrics, Tucson, USA). For image processing and karyotyping ColorProc, an in-house developed software tool, was used. A detailed protocol of the whole procedure has been published elsewhere (33).

Reverse transcriptase PCR analysis

Analysis of *H2AX* transcripts in $mCherry^+$ cells subjected to standard, *in trans* paired nicking and paired breaking gene editing, using either $gRNA^{H2AX.1}$ or $gRNA^{H2AX.2}$, was done as follows. Total RNA was extracted with the aid of the NucleoSpin RNA kit (Macherey-Nagel) essentially as specified by the manufacturer after adding 350 μl of RA1 buffer and 3.5 μl of β-mercaptoethanol (Merck). Reverse transcription on 1 μg of total RNA was performed at 50°C for 1 h with 200 ng of random primers, 0.2 mM dNTPs, 1× First-Strand Buffer, 5 mM dithiothreitol, and 200 U of SuperScript III Reverse Transcriptase (all from ThermoFisher Scientific). Next, 1-μl cDNA aliquots were subjected to PCR amplifications with the GoTaq G2 Flexi DNA Polymerase system (Promega; Cat. Nr.: M7808) using 0.4 μM of primer #1444 (5'-CAACGACGAGGAGCTCAACA-3'), 0.4 μM of primer #1508 (5'-GGCGGTGGTGGCCCTTAAAA-3'), 1 mM $MgCl_2$, 0.4 mM dNTPs, 1× GoTaq Flexi buffer and 1.25 U TaqI and H_2O to a final volume of 25 μl. Cycling parameters are specified in **Supplementary Table S16**. To serve as internal controls, 1-μl cDNA aliquots were also subjected to *GAPDH*-directed PCR amplifications with the GoTaq G2 Flexi DNA Polymerase system (Promega; Cat. No.: M7808) using, in this case, 0.4 μM of primer #119 (5'-AGCCACATCGCTCAGACACC-3') and 0.4 μM of primer #120 (5'-GTACTCAGCGCCAGCATCG-3'). Cycling parameters are specified in **Supplementary Table S16**. Finally, 10 μl PCR samples corresponding to *H2AX* and *GAPDH* transcripts were electrophoresed through a 2% (w/v) agarose gel in 1× TAE buffer.

Detection of indels by targeted amplicon sequencing

Target site genotyping of HeLa cell populations containing unmodified cells mixed with cells generated by gene editing involving standard, paired breaking or in trans paired nicking was performed as follows. PCR products spanning $gRNA^{H2AX.1}$ and $gRNA^{H2AX.2}$ target sites were amplified from total cellular DNA extracted from cells at two different timepoints by using the reagents and protocol provided in the DNeasy Blood & Tissue kit (Qiagen; Cat. No.: 69506). The cycling parameters and PCR mixture composition used for amplifying the *H2AX* target region are specified in **Supplementary Tables S16** and **S17**, respectively. *H2AX*-specific PCR products amplified from unmodified HeLa cell populations served as controls. Next, the amplicons corresponding to untagged *H2AX* alleles were extracted

following the QIAEX II Gel Extraction Kit (Qiagen Cat. No.: 20021) and were subjected to Sanger sequencing for determining indel frequencies and distributions with the aid of the ICE software https://ice.synthego.com/#/ (34).

Characterization of *PARP1* alleles in gene-edited cell populations

EGFP$^+$ HeLa cells resulting from *PARP1* gene tagging experiments using *in trans* paired nicking and standard gene editing protocols, were sorted with the aid of a BD FACSAria III flow cytometer (BD Biosciences). Next, total genomic DNA from these EGFP$^+$ populations and from unmodified HeLa cells was extracted by using the DNeasy Blood & Tissue Kit (Qiagen; Cat. No.: 69506), according to the manufacturer's instructions. The various DNA samples were subsequently subjected to PCR amplifications with two different primer pairs (i.e. primer pair A and B). Milli-Q water served as negative controls. The cycling parameters and PCR mixture compositions that were applied are indicated in **Supplementary Tables S16** and **S17**, respectively. Indels at *PARP1* alleles were detected by exposing amplicons to the mismatching-sensing T7EI (Biolabs) as below indicated. The presence of a 121-bp *PARP1* deletion in EGFP$^+$ HeLa cells generated through standard gene editing was established by direct Sanger sequencing of the low-molecular-weight species (241-bp) resulting from PCR with the primer pair B (**Supplementary Table S17**). Finally, the amplicons spanning the SpCas9-induced composite mutations were cloned using the TA cloning protocol (ThermoFisher Scientific Cat. No.: K1214) and were subsequently subjected to Sanger sequencing.

Identification and *in silico* analyses of *H2AX* and *OCT4* gRNAs

The number and distribution of candidate off-target sites for CRISPR complexes was probed by using publicly available algorithms (35,36). The UCSC Genome Browser (Assembly GRCh38/hg38) was used to display all canonical *S. pyogenes* CRISPR-SpCas9 gRNAs in and around the target sequences for tagging H2AX and OCT4. The tracks of the UCSC Genome Browser displayed in the present study are available through the links: https://genome.ucsc.edu/s/mafvg/hg38 Chen Tasca et al C-terminus H2AX CRISPR Zoom, https://genome.ucsc.edu/s/mafvg/hg38 Chen Tasca et al C-terminus CRISPR Zoom and https://genome.ucsc.edu/s/mafvg/hg38 Chen Tasca et al OCT4 CRISPR 1.5X. The computing of the predicted performance of each CRISPR-SpCas9 complex was made by a combination of algorithms in the crispor.org tool (36). The tracks for chained self-alignments and repeating elements are presented in full mode with the former depicting alignments of the human genome with itself after filtering out the redundant chromosomal positions that map to each other. As specified in the UCSC Genome Browser (Assembly GRCh38/hg38) website, the chained self-alignments and repeating elements tracks were generated with the aid of Blastz (37) and RepeatMasker (http://www.repeatmasker.org/), respectively.

Production and purification of lentiviral vector particles

The vesicular stomatitis virus glycoprotein G (VSV-G)-pseudotyped lentiviral vector LV.Cre was generated according to previously detailed protocols (38,39). In brief, 17 × 10^6 HEK293T cells were seeded per 175-cm^2 culture flask (Greiner Bio-One). The next day, the cells were transfected by adding to 19 ml of regular HEK293T cell culture medium, 1 ml of a 150 mM NaCl solution containing a mixture of 30 µg of DNA composed of lentiviral vector shuttle, packaging, and VSV-G-pseudotyping plasmids at a ratio of 2:1:1 (size-normalized for molecule copy number) and 90 µl of 1 mg ml^{-1} PEI solution (25 kDa PEI, Polysciences). The shuttle, packaging and pseudotyping constructs used were, BC17_pLV.Cre (Supplementary Information), psPAX2 (Addgene #12260) and pLP/VSVG (Invitrogen). The HEK293T cells were incubated overnight in a total 20-ml transfection mixture, after which, this transfection medium was removed and replaced by fresh DMEM supplemented with 5% FBS. At 3 days post-transfection, producer-cell conditioned media containing released vector particles were collected and the cellular debris were removed by centrifugation and filtration using 0.45-µm pore-sized HT Tuffryn membrane filter (Pall Life Sciences; Cat. No. PN4184). The resulting clarified supernatants were gently added onto 5-ml 20% (v/v) sucrose cushions in 35.8-ml polyallomer tubes (Beckman Coulter; Cat. No.: 326823). After ultracentrifugation (15,000 rpm for 2 h at 4°C) in an Optima LE-80K centrifuge (Beckman Coulter) using the SW28 rotor, vector-containing pellets were resuspended in 400 µl of ice-cold PBS pH 7.4 supplemented with 1% (w/v) bovine serum albumin. The vector particle titer of the purified LV.Cre stock was shown to be 31589 ng $p24^{gag}$ ml^{-1} after employing the RETROTEK HIV-1 p24 antigen ELISA kit reagents and protocol (ZeptoMetrix, Cat. No.: 0801111).

Quantification of *OCT4* gene targeting frequencies in iPSCs

Puromycin-resistant iPSCs resulting from *OCT4* gene editing via single nicking, *in trans* paired nicking, standard and paired breaking protocols, were seeded in wells of 24-well plates (Greiner Bio-One) at a density of 30 000 cells per well. The next day, LV.Cre was added to the iPSCs in a total volume of 500 µl of culture medium at a multiplicity-of-infection of 10 vector particles per cell. The frequency of iPSCs expressing OCT4::EGFP assembled via Cre-mediated recombination was determined by flow cytometry at 9 days and 18 days post-transduction.

Characterization of iPSCs with *OCT4* gene-edited alleles

Gene edited iPSCs expressing OCT4::EGFP after coupling *in trans* paired nicking to Cre-mediated recombination, were sorted through a BD FACSAria III flow cytometer (BD Biosciences) as single cell-deposited clones or as polyclonal populations. Both the OCT4::EGFP$^+$ clones and the OCT4::EGFP$^+$ cell populations were deposited in StemFlex Medium (ThermoFisher Scientific; Cat. No.: A3349401) containing 25 U ml^{-1} penicillin and 25 µg ml^{-1} of streptomycin (ThermoFisher Scientific) supplemented with Revitacell (ThermoFisher Scientific). The medium of the iPSC clones was replenished every other day. The medium was refreshed every day when the wells of 96-well plates (Greiner Bio-One) contained visible clusters of viable cells. These cell colonies were further expanded into wells of 48-well plates (Greiner Bio-One) and subsequently into wells of 24-well plates (Greiner Bio-One). Finally, they were expanded and adapted to grow in E8 medium. The OCT4::EGFP$^+$ iPSC clones and iPSC polyclonal populations were subsequently subjected to OCT4/EGFP dual-colour confocal microscopy and flow cytometry assays. Finally, the pluripotency of iPSCs was assessed after applying differentiation protocols and confocal microscopy analyses as detailed under the section 'Differentiation of iPSCs'.

Confocal microscopy analyses

Cells seeded in glass coverslips were fixed in 2% or 4% (v/v) paraformaldehyde (PFA) and were permeabilized in 0.5% (w/v) Triton X-100 in Tris-buffered saline (TBS) pH 7.6 (50 mM Tris–HCl pH 7.6, 150 mM NaCl) at room temperature for 5–10 min (**Supplementary Table S18**). Subsequently, the cells were incubated for 1 h to 2 h with blocking Antibody Diluting Solution (Abdil) consisting of TBS, Triton X-100, 2% (w/v) bovine serum albumin and 0.1% sodium azide. In-between each fixation, permeabilization and blocking steps, the specimens were washed three times for 5 min at room temperature with 0.1% Triton X-100 in TBS. The primary antibodies were diluted in Abdil (**Supplementary Table S18**) and were added to the cells for 1 h at room temperature. After three 5-min washes with 0.1% Triton X-100 in TBS, the cells were incubated with fluorochrome-conjugated secondary antibodies diluted in Abdil for 30 min to 1 h in the dark at room temperature (**Supplementary Table S18**). Next, the specimens were subjected to three 5-min washes with 0.1% Triton X-100 in TBS and were mounted in ProLong Gold Antifade Mounting reagent containing DAPI (ThermoFisher Scientific; Cat. No.: P36931) or in VECTASHIELD Antifade Mounting Medium (VECTOR; Cat. No.: H-1000). Before the addition of the latter mounting medium, the specimens were incubated for 5 min in the dark with the DNA staining reagent DAPI (Invitrogen Cat. No.: R37606) diluted 1:1000 in TBS. Finally, fluorescence microscopy was carried out with an upright Leica SP8 confocal microscope (Leica Microsystems) equipped with Leica hybrid detectors, HyD (Leica Microsystems).

Differentiation of iPSCs

The culturing of clumps of iPSCs on glass coverslips coated with VTN-N triggered the spontaneous differentiation of iPSCs along the three embryonic germ layers. In brief, iPSCs were treated with PBS-EDTA for 1 min at 37°C and were subsequently gently dissociated into large cell clumps by scrapping. The resulting cell clumps were then cultured in suspension for 24 h on low-attachment plates at 37°C. Next, the iPSCs were seeded on coverslips coated with VTN-N in Essential 8 medium (ThermoFisher Scientific, #A1517001) supplemented with Revitacell (ThermoFisher Scientific, Cat. #A2644501). The day after, the medium was changed to DMEM/F12 growth medium (Gibco Cat. #31331–028) containing 20% FBS (Biowest Cat. #S1860–500). The DMEM/F12 medium was replenished every 2–3 days. After 3 weeks under differentiation conditions, the iPSCs were processed for immunofluorescence confocal microscopy for the detection of markers characteristic of the endoderm, mesoderm and ectoderm lineages (**Supplementary Table S19**). The markers corresponding to the three embryonic germ layers that were tested were α-fetoprotein (AFP), forkhead box protein A2 (FOXA2), α-smooth muscle actin (α-SMA), endothelial cell adhesion molecule-1 (CD31), and tubulin β3 class III (TUBB3).

T7 endonuclease I-based genotyping assays

Genotyping assays based on the mismatch-sensing T7EI enzyme were performed for detecting indels at target sequences of CRISPR complexes located at human *PARP1*, *RAG1* and *AAVS1* alleles and at off-target chromosomal positions located in the human *OCT4* pseudogenes *POU5F1P4* and *POU5F1P5*. For the latter assays, the genomic DNA of puromycin-resistant iPSC populations grown after *OCT4*-targeting experiments was extracted by using the DNeasy Blood & Tissue Kit and protocol (Qiagen, Cat. No.: 69506). The GoTaq G2 Flexi DNA Polymerase system (Promega; Cat. No.: M7808) was subsequently applied to amplify the *POU5F1P4* and *POU5F1P5* genomic sequences. The cycling parameters and PCR mixture compositions are specified in **Supplementary Tables S16** and **S17**, respectively. Next, the resulting amplicons were subjected to the thermocycling procedure indicated in **Supplementary Table S20** after which, 10-µl samples were incubated at 37 °C for 17 min with 1.5 µl 10× NEBuffer 2, 0.5 µl (5U) T7EI (New England Biolabs; Cat. No.: M0302) and 3 µl of Milli-Q water. Samples that were not treated with T7EI provided for negative controls. Finally, after agarose gel electrophoresis, untreated and T7EI-treated amplicons were analysed by using the Gel-Doc XR+ system and the ImageLab 4.1 software (both from Bio-Rad).

Flow cytometry

The frequencies of cells expressing H2AX::mCherry, EGFP::PARP1, OCT4::EGFP and EGFP were determined by using a BD LSR II flow cytometer (BD Biosciences). Parental unmodified cells or cells corresponding to experimental negative controls were used to establish the thresholds corresponding to background fluorescence. At least 10 000 viable single cells were analysed per sample. Data were analysed with the aid of FlowJo 10.5.0 software (Tree Star).

Western blotting

After two washes with ice-cold PBS pH 7.4, sorted EGFP::PARP1^{+} and EGFP::PARP1^{-} HeLa cells that had been exposed to standard gene editing or in trans paired nicking procedures were collected from wells of six-well plates by adding 250 µl of lysis RIPA buffer (Pierce Cat. No.: 89900) supplemented with a protease inhibitor cocktail (cOmplete Mini, Sigma-Aldrich Cat. No.: 11836153001). Untreated HeLa cells were taken along as negative controls. The cell lysates were subsequently passed thrice through a 1 ml syringe with a 26 GA 3/8 0.45 × 10 needle (BD Plastipak Cat. No.: 300015) and spun at 14 000 RPM for 5 min at 4°C in an Eppendorf 5424 centrifuge. The protein concentrations in the resulting supernatants were determined by using the BCA Protein Assay Kit (ThermoFisher Scientific Cat. No.: 23225) according to the manufacturer's instructions. Next, 15 µg of protein were diluted in 4× sample buffer and 20× reducing agent (both from Bio-Rad Cat. No. 161-0791 and 161-0792, respectively) and incubated at 95°C for 5 min. Protein samples were loaded in a 7% SDS-PAGE gel. After electrophoreses, the proteins were transferred to a PVDF membrane (Millipore Immobilon Cat. No.: IPVH00010) and were blocked overnight in TBS with 0.05% (v/v) Tween 20 (TBST, ThermoFisher Scientifc Cat. No.: 28358) supplemented with 5% (w/v) Elk milk (Campina). Next, the membrane was incubated with PARP1 polyclonal antibody (Thermo Fisher, Cat. No.: PA5-34803) diluted 1:5,000 in blocking buffer or with α/β tubulin antibody (Cell Signalling Cat. No.: CST 2148) diluted 1:5000 in blocking buffer. After an overnight incubation period at 4°C, the membranes were washed in TBST and incubated for 4 h at 4°C with an anti-rabbit IgG secondary antibody conjugated to horseradish peroxidase (IgG-HRP; Santa Cruz Cat. No.: sc-2004) diluted 1:1,000 in TBST. Proteins were detected by using horseradish peroxidase substrate Pierce ECL2 (Pierce Cat. No.: 80196) following the manufacturer's specifications and Super RX-N X-ray film (Fujifilm).

Statistical analyses

With the exception of genomic DNA samples used for assessing genome-wide off-target effects of CRISPR complexes by orthogonal HTGTS analyses, the researchers were not blinded to sample allocation. Statistical analyses were performed on data sets derived from a minimum of three biological replicates done on different days. These data were analyzed by using the GraphPad Prism 8.0.1 software. The statistical significances were calculated with the tests indicated in the figure legends. P values lower than 0.05 were considered to be statistically significant.

RESULTS

Distinct prevalence of genome-wide rearrangements after Cas9 versus Cas9^{D10A} delivery

Genome-wide off-target effects of programmable nucleases are commonly assessed by high-throughput sequencing of exogenous DNA tags 'trapped' at two-ended DSB termini or, more recently, in situ detection of DSB repair factors (40,41). Although SSBs are mostly resolved through conservative repair processes they can in principle lead to DSBs if a replication fork advances through them and collapses (42). However, the resulting one-ended chromosomal breaks are unlikely substrates for exogenous DNA 'trapping'. Therefore, to fulfil the lack of a sensitive and unbiased genome-wide assay for comparing off-target effects triggered by programmable nucleases versus programmable nickases, we have adapted the HTGTS assay (10). In contrast to other approaches, HTGTS detects off-target effects by deep sequencing of translocations joining bait and prey DSBs made by universal and test nucleases, respectively (**Figure 1A**). In addition to taking place at *bona fide* target sites, prey DSBs can also occur at off-target sites of a specific test nuclease under examination. In adapting the HTGTS assay for comparing off-target effects induced by nucleases versus nickases, we assured that bait DSBs are exclusively made by a universal nuclease whilst prey DSBs are instead generated by either test nucleases or test nickases. To this end, we combined *S. pyogenes* SpCas9 with its ortholog *Staphylococcus aureus* Cas9 (SaCas9). In particular, test *S. pyogenes* and universal *S. aureus* CRISPR complexes were designed for generating prey DNA lesions (i.e. SSBs or DSBs) and universal bait DSBs, respectively (orthogonal HTGTS). After selecting RAG1-targeting SaCas9:Sa-gRAG1.1 complexes as inducers of bait DSBs (**Supplementary Figure S1**), HEK293T cells were exposed to these complexes together with SpCas9:gAAVS1 or SpCas9^{D10A}:gAAVS1, each cleaving or nicking, respectively, at the commonly used *AAVS1* safe-harbour locus (**Figure 1B**). As expected, genotyping assays based on the mismatch-sensing T7EI enzyme, readily revealed indels at *RAG1* and *AAVS1* in cells subjected to SaCas9:Sa-gRAG1.1 and SpCas9:gAAVS1 (**Supplementary Figure S2**). In contrast, indels were detected at *RAG1* but not at *AAVS1* in cells treated with SaCas9:Sa-gRAG1.1 and SpCas9^{D10A}:gAAVS1, confirming that the latter complex displays low mutagenicity at the target intron (**Supplementary Figure S2**) (12). Control orthogonal HTGTS read libraries generated by delivering SaCas9:Sa-gRAG1.1 alone, besides detecting a single poorly-enriched off-target site on chromosome 1, revealed a genome-wide translocation pattern consistent with previously described *S. pyogenes* SpCas9:gRNA bait libraries (**Figure 1C**, **Supplementary Figures S3** and **S4**) (10). Importantly, applying orthogonal HTGTS analyses to experimental DNA samples (**Figure 1C**, **Supplementary Figures S3** and **S4**), demonstrated that amidst cells exposed to SpCas9:gAAVS1 and SpCas9^{D10A}:gAAVS1, the former had significantly higher numbers of off-target translocation hotspots than the latter; i.e. 30.7 ± 6.4 versus 0.7 ± 0.6 recurrent hotspots, respectively (**Figure 1C** and **D** and **Supplementary Figure S4**). In addition, SpCas9:gAAVS1 yielded higher frequencies of translocation junctions per hotspot than SpCas9^{D10A}:gAAVS1 (**Figure 1E** and **Supplementary Figure S4**). It is also noteworthy that, amongst the two translocation hotspots associated with SpCas9^{D10A}:gAAVS1 activity, was that involving *RAG1* bait and *AAVS1* prey target DNA (**Figure 1C**, **Supplementary Figures S3** and **S4**). This data suggests that individual SSBs can indeed be processed into chromosomal DSBs in living mammalian cells.

Together, these data establish orthogonal HTGTS as a sensitive method for the unbiased genome-wide detection of off-target effects elicited by genomic SSBs. Importantly, these results also lend support to SpCas9^{D10A} as a genome-editing tool that diminishes allelic and non-allelic chromosomal mutations and rearrangements.

***In trans* paired nicking minimizes disruptive genotype-phenotype associations**

Earlier *AAVS1* gene targeting experiments in HeLa cells and human induced pluripotent stem cells (iPSCs) demonstrated that DSB-dependent gene editing approaches yield more inaccurate and random donor DNA insertions than *in trans* paired nicking (12). Besides augmenting genotype-phenotype unpredictability (e.g. via insertional mutagenesis), random chromosomal DNA integration results in transgene expression variegation due to chromosomal positional effects (11-12). Similar *AAVS1* gene targeting experiments performed in HEK293T cells support these previous findings (11-12) by showing that heterogeneous transgene expression is prevalent in cell populations subjected to donor plasmids and DSB-forming nucleases (**Supplementary Figure S5**).

Tagging endogenous proteins with fluorescent reporters is a frequent goal of genome editing endeavours, including for establishing live-cell screening systems or studying cellular processes in a spatiotemporal fashion. However, the need for targeting gene termini limits the availability of gRNAs with potentially high activities and/or specificities. The presence of functional motifs further limits gRNA

design as, often, HDR-mediated knock-in of one allele is accompanied by NHEJ-induced knockout of the other allele creating functional gene-dose imbalances. The gRNA availability issue becomes extreme in cases where target sequences (coding or otherwise) are not unique in the genome. These sequences are in fact dubbed 'impossible to target' in the CRISPR tracks of the UCSC Genome Browser and are defined as having at least one identical copy in the genome (43). Thus, as challenging targets for comparing the performance of SpCas9 versus SpCas9^{D10A}, we sought to tag housekeeping *H2AX* and cell type-specific *OCT4* alleles with live-cell reporters. The difficulty in tagging these genes stems from the fact that H2AX function depends on a C-terminal SQ phosphorylation motif (44) that restricts gRNA selection in this coding region and *OCT4* termini share 100% sequence identity with sequences found in four autosomal pseudogenes that prevents the identification of *OCT4*-specific gRNAs.

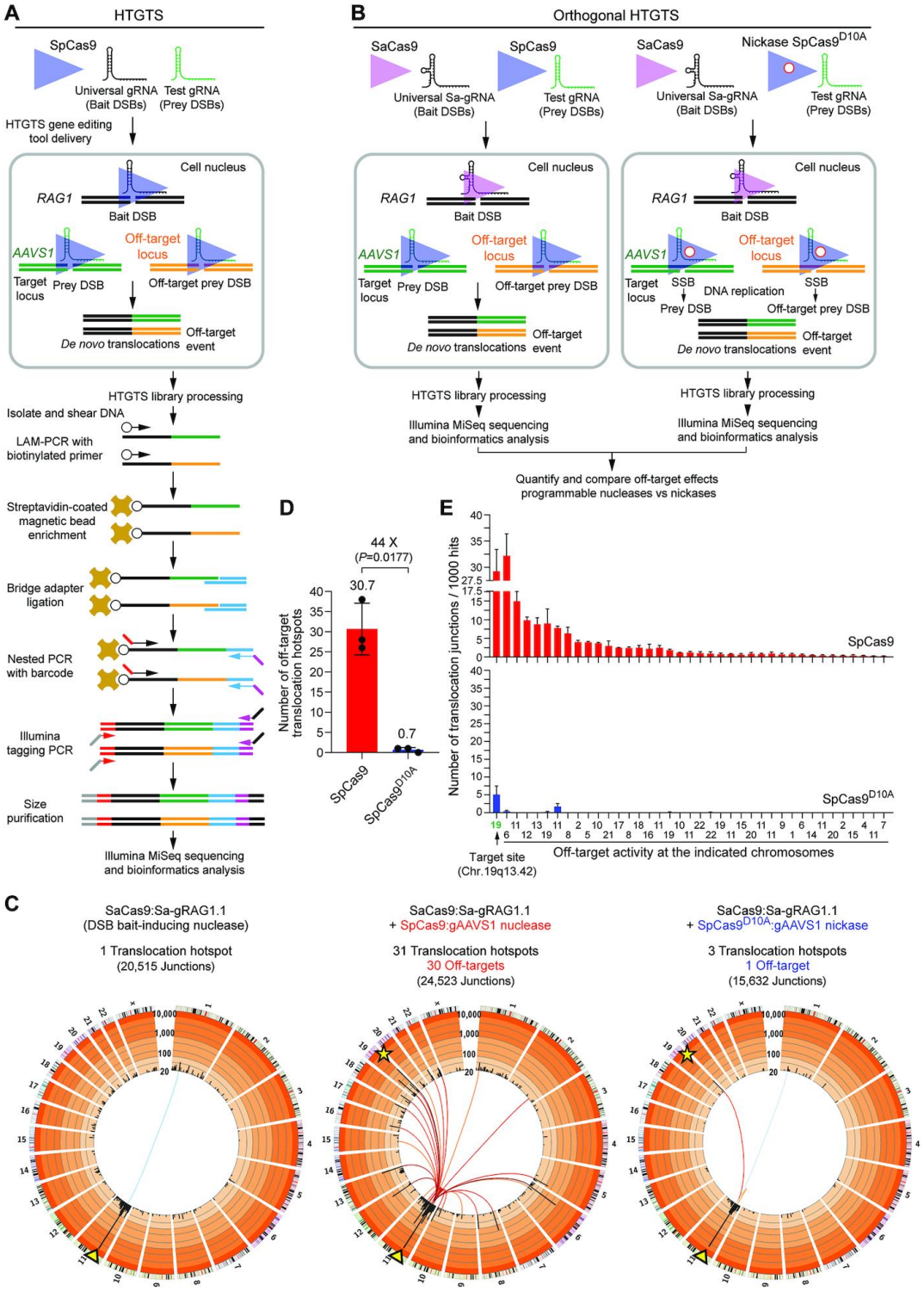


Figure 1. Comparing off-target effects triggered by cleaving versus nicking CRISPR complexes. (**A**) Diagram of the HTGTS pipeline for detecting SpCas9-induced off-target effects. Cells are exposed to *S. pyogenes* CRISPR complexes containing universal and test gRNAs that induce bait and prey DSBs at *RAG1* and target loci, respectively. The prevalence and distribution of off-target hotspots conferred by test gRNAs are determined by an HTGTS pipeline comprising next-generation sequencing of translocations between *RAG1* and off-target DNA (black and orange lines, respectively). (**B**) Diagram of the orthogonal HTGTS pipeline for detecting SpCas9^{D10A}-induced off-target effects. Orthogonal HTGTS assays make use of *S. aureus* and *S. pyogenes* CRISPR complexes for generating bait DSBs at *RAG1* and either prey DSBs or nicks at target loci, respectively. The orthogonality (i.e. lack of cross-talk) between gRNAs and Cas9 proteins from these CRISPR systems avoids nicking at *RAG1* and cleaving at off-target sites of test SpCas9^{D10A}:gRNA complexes (right panel). Further, exchanging SpCas9^{D10A} by SpCas9 in parallel orthogonal HTGTS assays permits comparing side-by-side genomic disruptions inflicted by cleaving versus nicking CRISPR complexes (left panel). Original and orthogonal HTGTS assays share the same downstream library processing and bioinformatics analysis steps. (**C**) Cumulative orthogonal HTGTS analyses (i.e. Circos plots) from three biological replicates. Arrowheads on chromosome 11 indicate the location of the SaCas9:Sa-gRAG1.1 universal bait DSB for all sequence read libraries; stars on chromosome 19 mark the AAVS1 target site of test *S. pyogenes* CRISPR complexes. Blue-graded lines from bait DSBs at the *RAG1* locus indicate bait-related off-targets whereas red-graded lines indicate test gAAVS1-related translocation hotspots from the activity of S. pyogenes CRISPR complexes at target and off-target sites. Hotspots are established only when significantly enriched translocation sites are present in the majority of independent HTGTS replicate experiments (n≥2). Black bars represent 5 Mb bins across each chromosome and enrichment is displayed on a custom color coded log scale by order of magnitude. (**D**) Number of gAAVS1 off-target translocation hotspots in SpCas9 and SpCas9^{D10A} sequence read libraries. Significance was calculated with paired two-tailed Student's *t* tests. (**E**) Relative frequencies of junctions per gAAVS1 translocation hotspot in SpCas9 and SpCas9^{D10A} sequence read libraries. Individual experimental values and respective Circos plots are shown in **Supplementary Figures S3** and **S4**, respectively. Bars and error bars in panels **D** and **E** indicate mean ± S.D., respectively (n = 3 independent biological replicates).

H2AX gene editing experiments were initiated by transfecting HeLa cells with plasmids expressing cleaving SpCas9:gRNA or nicking SpCas9^{D10A}:gRNA complexes containing gRNA$^{H2AX.1}$ or gRNA$^{H2AX.2}$ (**Figure 2A**). The transfection mixtures included donor constructs pDonor^{H2AX} or pDonor$^{H2AX.TS}$. The latter differs from the former in that it has the *H2AX*-specific gRNA target sites flanking the targeting module consisting of 'homology arms' and a mCherry reporter tag (**Figure 2A** and **B**). After delivering these tools, we sought to access the efficiency and precision of gene editing involving (i) DSBs on target DNA (standard), (ii) DSBs on target and donor DNA (paired breaking; DSB^2), (iii) SSBs on target DNA (single nicking) and (iv) SSBs on target and donor DNA (*in trans* paired nicking; Nick^2) (**Figure 2B**). The efficiency and precision of *H2AX* gene editing was ascertained by combining flow cytometric quantification of mCherry$^+$ cells with molecular analysis of randomly isolated mCherry$^+$ clones, each of which, representing an individual genome-modifying event. Importantly, we exploited the fact that the mCherry-tagged intronless *H2AX* gene in donor plasmids behaves as an autonomous reporter unit (**Figure 2C**, top panel) to avoid biased selection of cells harbouring targeted exogenous DNA chromosomal insertions (targeted integrants). The frequencies of transiently and stably transfected cells were determined by flow cytometry before and after episomal templates had been eliminated through sub-culturing (**Figure 2C**, top and bottom panel, respectively). This analysis revealed that, for both gRNAs used, *in trans* paired nicking yielded ~4-fold higher percentages of stably transfected cells than those resulting from the single nicking approach (**Figure 2C**, bottom panel). The robust enhancement on the frequencies of genetically modified cells achieved by *in trans* paired nicking over those resulting from the single nicking strategy is consistent with previous experiments targeting introns (12). Hence, in addition to supporting initial theoretical models postulating nicked DNA partners as homologous recombination substrates (45), these results further stress the limited utility of the single nicking approach. The paired breaking strategy led to the highest frequencies of stably transfected cells (**Figure 2C**, bottom panel). However, it is worth noting that the attendant free-ended donor DNA templates created in cellula by SpCas9-mediated cutting (paired breaking) are prone to yielding complex genome-modifying events, i.e., off-target and inaccurately targeted chromosomal insertions, including concatemeric and HDR-independent integrants (2,11–13). Indeed, although genetically modified cells expressed tagged *H2AX* transcripts independently of the gene editing procedure used (**Supplementary Figure S6**), junction PCR screens of randomly selected mCherry$^+$ clones readily revealed that paired breaking yielded the least precisely targeted integrants when compared to standard and *in trans* paired nicking (**Figure 2D** and **Supplementary Figure S7**). Notably, untagged *H2AX* alleles in mCherry$^+$ clones exposed to SpCas9 and SpCas9^{D10A} had varying and uniform sizes, respectively (**Supplementary Figure S7**). These results support recent findings indicating that, in addition to short indels, SpCas9 can induce gross structural variants at target sequences, such as, large insertions and deletions (4,10). To further characterize these collateral gene-editing events, nucleotide sequencing of *H2AX* alleles was done in mCherry$^+$ clones modified through either standard or *in trans* paired nicking procedures. This target site genotyping analysis confirmed the presence of a range of indel footprints in mCherry$^+$ cells obtained via standard gene editing (**Supplementary Figure S8**). In contrast, untagged

H2AX alleles remained intact in mCherry^{+} cells generated through *in trans* paired nicking (**Supplementary Figure S8**), with the respective tagged *H2AX* alleles expressing the H2AX::mCherry fusion protein in the nuclei of the respective cell populations (**Figure 2E**).

For further assessing the accuracy and mutagenicity of the different gene editing strategies (**Figure 2D** and **Supplementary Figure S7**, respectively), we randomly selected mCherry^{+} clones from cultures initially exposed to the gRNA with the fewest predicted off-target sites, i.e., gRNA$^{H2AX.2}$ (**Supplementary Figure S9**). Interestingly, gRNA$^{H2AX.2}$ directs SpCas9 and SpCas9^{D10A} to cut and nick, respectively, within the codons of the previously mentioned SQ phosphorylation motif whose integrity is crucial for H2AX function (**Figures 2A** and **Supplementary Figure S9**). In this regard, it is worth noting that reduced H2ax dosages in heterozygous *H2ax*$^{+/-}$ knockout mice have uncovered pleiotropic haploinsufficiency phenotypes (46). For instance, embryonic fibroblasts from these *H2ax*$^{+/-}$ mice present growth kinetic curves that are in-between those of wild type and homozygous *H2ax*$^{-/-}$ mice (46). Thus, we next compared the fitness of human cells whose *H2AX* loci had been edited by either *in trans* paired nicking or DSB-dependent gene editing approaches. To this end, populations of mCherry^{+} cells were mixed with a small fraction of unmodified cells (i.e. 5%) and were subsequently monitored by flow cytometry upon serial sub-culturing rounds. Such cell competition settings demonstrated a fitness loss (i.e. growth disadvantage) specifically in cells that had undergone standard and paired breaking gene editing after SpCas9:gRNA$^{H2AX.2}$ delivery (**Figure 2F**). This loss-of-fitness phenotype correlated with the time-dependent disappearance of cells harboring *H2AX* indels disabling the SQ phosphorylation target motif (**Supplementary Figure S10**). We also performed competition experiments in which edited cells had initially been exposed to gRNA$^{H2AX.1}$ instead of gRNA$^{H2AX.2}$. Although displaying a higher potential for off-target effects than gRNA$^{H2AX.2}$, gRNA$^{H2AX.1}$ has a lower change of disrupting the SQ protein motif (**Figure 2A** and **Supplementary Figure S9**). In this case, we observed neither the replacement of edited cells by unedited cells (**Supplementary Figure S11**) nor the elimination of cells with DSB-derived *H2AX* indels (**Supplementary Figure S12**). Thus, in contrast to a process of 'purification' from mutations at the cost of gene-edited cell loss, there was instead, gene-edited cell maintenance at the cost of a 'fixation' of mutations in the populations subjected to SpCas9:gRNA$^{H2AX.1}$ complexes (**Supplementary Figures S11** and **S12**). Importantly, reminiscent of the previous sequencing of *H2AX* alleles in individual clones (**Supplementary Figure S8**), the population-level *H2AX* genotyping assays further confirmed the non-disruptive character of *in trans* paired nicking by revealing the striking dominance of gene edited cells lacking *H2AX* mutations at both time points analysed, independently of the gRNA used (**Supplementary Figures S10** and **S12**, bottom D panels). Taken together, these data indicate that the loss-of-fitness phenotype seen in SpCas9:gRNA$^{H2AX.2}$-treated cells (**Figure 2F**) is attributable to functional H2AX haploinsufficiency caused by NHEJ-mediated disruption of the SQ post-translational modification motif (**Supplementary Figures S8** and **S10**).

***In trans* paired nicking minimizes mutagenesis within coding sequences of target alleles**

PARP1, like H2AX, is also involved in DNA repair, however, functional redundancies with other PARP family members are reported (46,47). Tagging PARP1 with EGFP after delivering conventional pDonor^{PARP1} or target site-containing pDonor$^{PARP1.TS}$, together with cleaving Cas9:gRNA^{PARP1} or nicking Cas9^{D10A}:gRNA^{PARP1} complexes (**Figure 3A**), revealed that *in trans* paired nicking and standard gene editing led to higher frequencies of stably transfected cells than those reached by using the single nicking approach (**Figure 3B**). Importantly, junction PCR screens of randomly isolated EGFP^{+} clones confirmed accurate DNA targeting events in cell populations subjected to *in trans* paired nicking and standard gene editing (**Figure 3C**). Moreover, cell competition experiments involving tracking mixtures of unedited and *PARP1*-edited cells provided no evidence for cell-fitness losses in each of the EGFP::PARP1-expressing populations (**Figure 3D**). Despite this, we sought to characterize EGFP::PARP1^{+} and EGFP::PARP1^{-} cell populations obtained through in trans paired nicking versus standard gene editing (**Figure 4A**). In addition to the typical small indels established after NHEJ-mediated DSB repair, the EGFP::PARP1^{+} cell fraction generated through standard gene editing contained large *PARP1* deletions (**Figures 4B** and **4C**). Of note, small indels were even detected in the EGFP::PARP1^{-} cell fraction isolated from cultures subjected to standard gene editing (**Figure 4C**). Sequence analysis of *PARP1* target DNA in EGFP::PARP1^{+} cells identified a 121-bp deletion mixed with shorter deletions of varying sizes (**Figure 4D** and **Figure 4E**, respectively). These structural variants are reminiscent of those detected in mCherry^{+} cells that had been exposed to cleaving *H2AX*-specific CRISPR-Cas9 complexes (**Supplementary Figures S7** and **S8**), and further support the data indicating

that targeted DSBs can trigger gross genomic alterations (4,10). In contrast, *PARP1* structural variants consisting of large deletions and small indels were detected neither in EGFP::PARP1^{+} nor EGFP::PARP1^{-} cell fractions generated after *in trans* paired nicking (**Figures 4B-D**).

Finally, dual-colour confocal microscopy showed that, regardless of the gene editing methodology, EGFP-tagged PARP1 localized properly in cell nuclei (**Figure 5A**). Tellingly, however, western blot analysis revealed that contrary to EGFP::PARP1^{+} cells resulting from *in trans* paired nicking, EGFP::PARP1^{+} cells derived from standard gene editing suffered a substantial depletion of the endogenous, untagged, PARP1 protein (**Figure 5B**). This data is consistent with the high prevalence of *PARP1* structural variants in EGFP::PARP1^{+} cells initially treated with pDonor^{PARP1} and Cas9:gRNA^{PARP1} (**Figures 4B-E**).

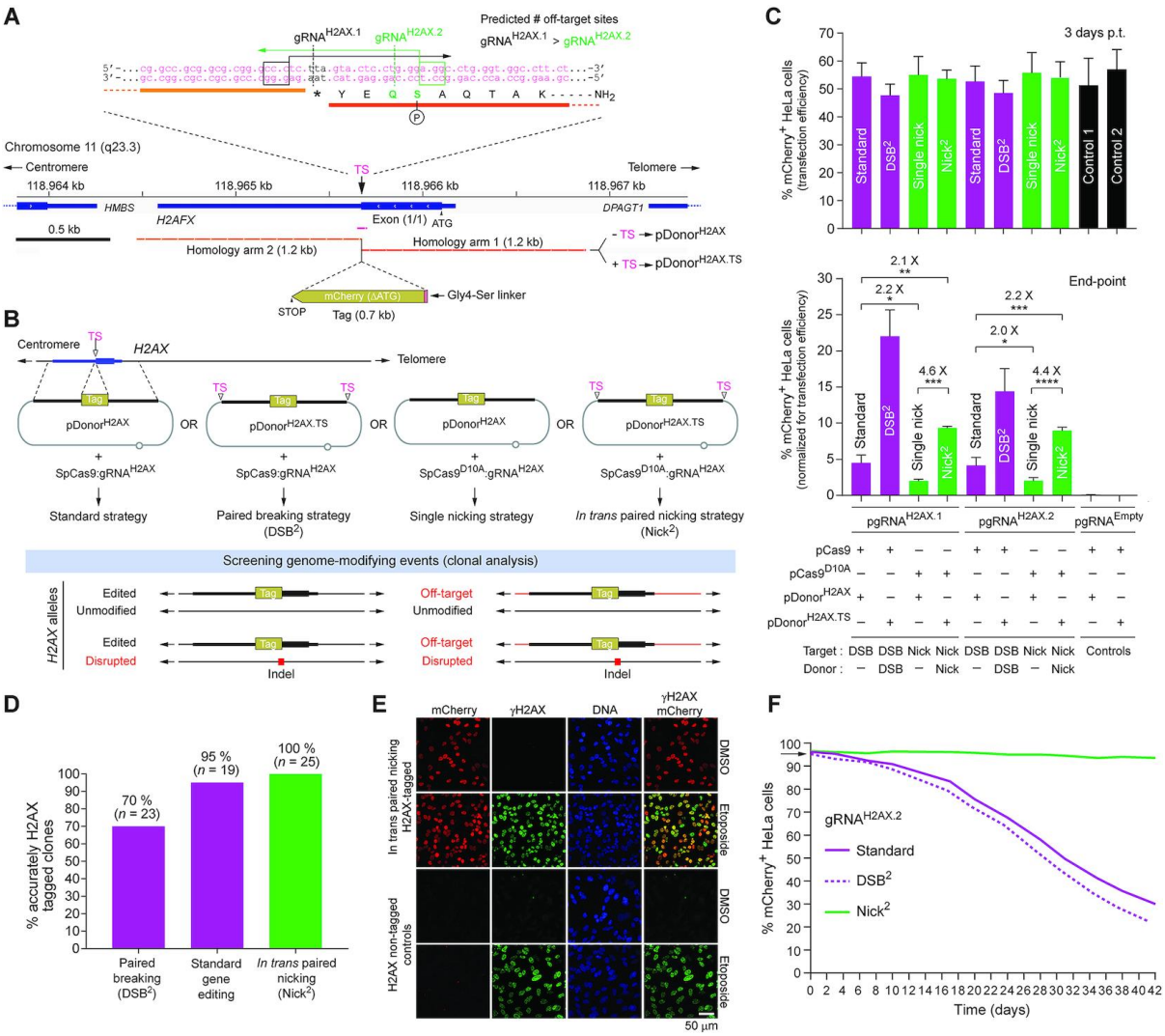


Figure 2. Homology-directed *H2AX* gene editing based on cleaving or nicking CRISPR complexes. (**A**) Diagram of the *H2AX* genomic region. The gRNA$^{H2AX.1}$ and gRNA$^{H2AX.2}$ target sites (TS) are highlighted by horizontal arrows and boxed nucleotides (PAMs). The H2AX post-translationally phosphorylated serine residue 129 is marked with a circled P. The donor plasmids pDonor^{H2AX} and pDonor$^{H2AX.TS}$ contain as targeting module *H2AX* sequences ('arms of homology') flanking a *mCherry* tag. (**B**) Schematics of *H2AX* gene editing strategies. Standard and paired breaking gene editing involve DSB formation at the genomic TS or at this TS and those in the donor DNA, respectively. Single nicking and *in trans* paired nicking gene editing comprise SSB formation at the genomic TS or at this TS and those in the donor DNA, respectively. Wanted and unwanted (red icons) genome-modifying events are depicted. (**C**) Quantification of transiently and stably transfected human cells. Flow cytometry was done on HeLa cell cultures co-transfected with the indicated plasmids. Top and bottom graphs, frequencies of mCherry^{+} cells at early and late time points after transfection (3 days and 2 weeks, respectively). Data are presented as mean ± s.e.m. of four independent biological replicates. Significance between the indicated datasets was calculated with one-way ANOVA followed by Tukey's test for multiple comparisons; *P< 0.05; **P<0.01; ***P< 0.001; ****P< 0.0001. (**D**) Assessing *H2AX* gene editing accuracy. The frequencies of precisely targeted mCherry+ clones were determined through junction PCR screens (**Supplementary Figure S7**). (**E**) Confocal microscopy analysis of *H2AX* gene-edited cells. HeLa cells genetically modified by *in trans* paired nicking were

subjected to direct and indirect fluorescence microscopies for detecting, respectively, mCherry and H2AX phosphorylated at Ser-126 (γH2AX). Prior to microscopy, the cells were incubated with a DNA damaging antitumor agent (etoposide) or with vehicle (DMSO). Nuclei were stained with DAPI. (**F**) Competition experiments comprising unedited and H2AX edited cells. Long-term cultures of cells expressing H2AX::mCherry (95% at t = 0 days) mixed with unedited cells (5% at t = 0 days) were monitored by flow cytometry. H2AX tagging was done through standard, paired breaking (DSB_2), or in trans paired nicking ($Nick_2$) gene editing using $gRNA^{H2AX.2}$.

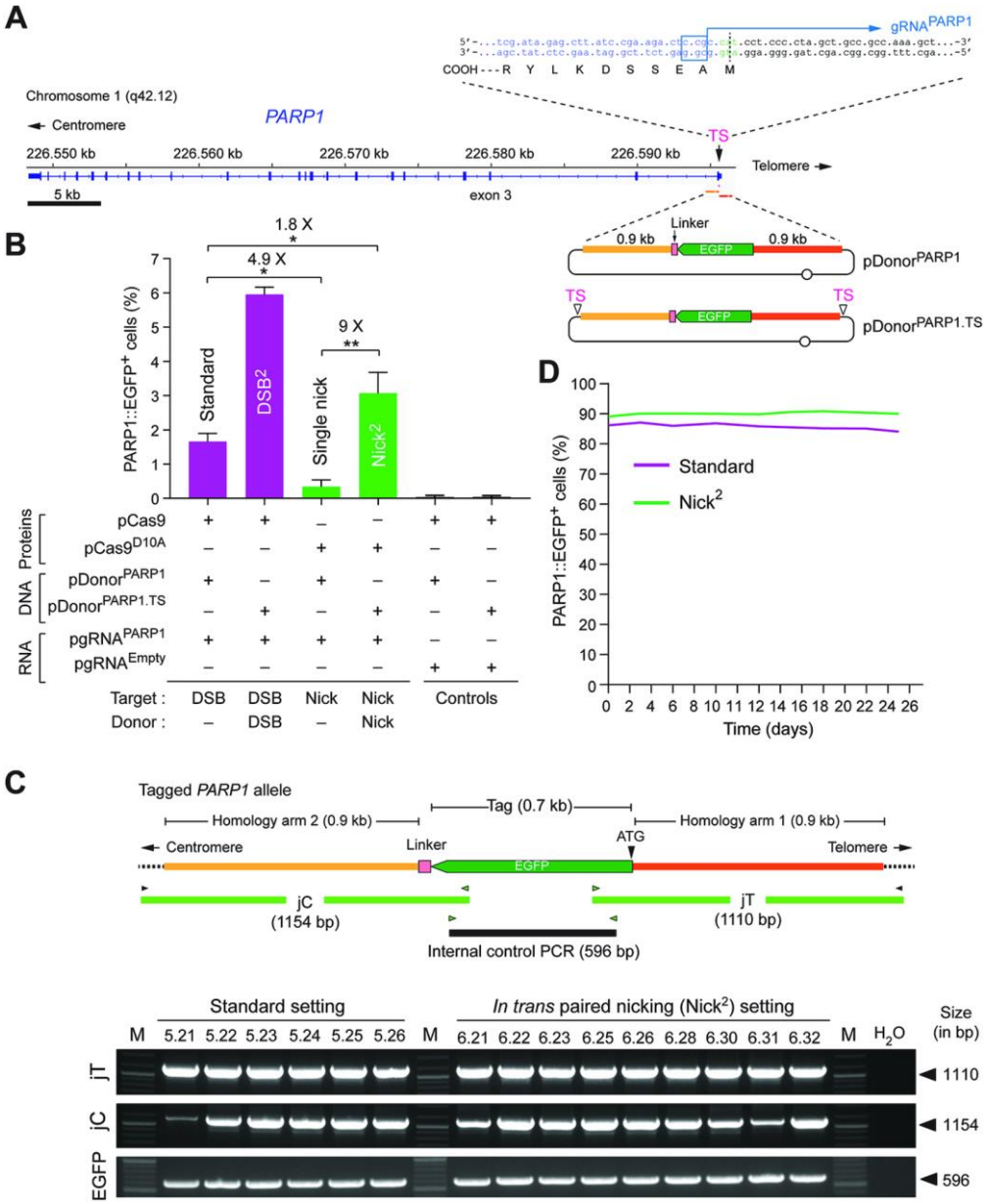


Figure 3. Homology-directed *PARP1* gene editing based on cleaving or nicking CRISPR complexes. (**A**) Diagrams of *PARP1* and *PARP1*-tailored gene editing tools. The gRNAPARP1 target site (TS) is indicated by the horizontal arrow and boxed nucleotides (PAM). The vertical dashed line marks the SpCas9:$gRNA^{PARP1}$ cleaving position. The N-terminal PARP1 amino acids are drawn next to their respective codons. The donor constructs $pDonor^{PARP1}$ and $pDonor^{PARP1.TS}$ have as targeting module *PARP1* sequences ('arms of homology') flanking a *EGFP* tag. The latter construct has, in addition, TS sequences flanking the targeting module. (**B**) Quantification of genetically modified human cells. Flow cytometry of HeLa cell cultures co-transfected with the indicated plasmids. Data are presented as mean ± S.D. of three independent biological replicates. Significance between the indicated datasets was calculated with one-way ANOVA followed by Tukey's test for multiple comparisons; **P*< 0.05; ***P*< 0.01. (C) Molecular characterization of human cells genetically modified through standard versus in trans paired nicking gene editing at *PARP1*. Top panel, Junction PCR assay for assessing *PARP1* gene tagging. Amplicons diagnostic for HDR-derived centromeric and telomeric junctions between exogenous DNA and *PARP1* (jC and jT, respectively) are depicted. Amplicons specific for *EGFP* served as internal controls (EGFP). Bottom panel, Junction PCR analysis on genomic DNA from $EGFP^+$ HeLa cell clones retrieved from cultures co-transfected with pCas9, $pDonor^{PARP1}$ and $pgRNA^{PARP1}$ (Standard setting) or with $pCas9^{D10A}$, $pDonor^{PARP1.TS}$ and $pgRNA^{PARP1}$ (*In trans* paired nicking setting). H_2O, PCR sample containing nuclease-free water instead of genomic DNA. Lanes M, GeneRuler DNA Ladder Mix molecular weight marker. (D) Competition experiment involving unedited and *PARP1* edited cells. Long-term cultures of HeLa cells expressing EGFP-tagged PARP1 mixed with unedited cells were monitored by flow cytometry. Green and magenta lines, $EGFP^+$ cells generated by *in trans* paired nicking and standard gene editing, respectively.

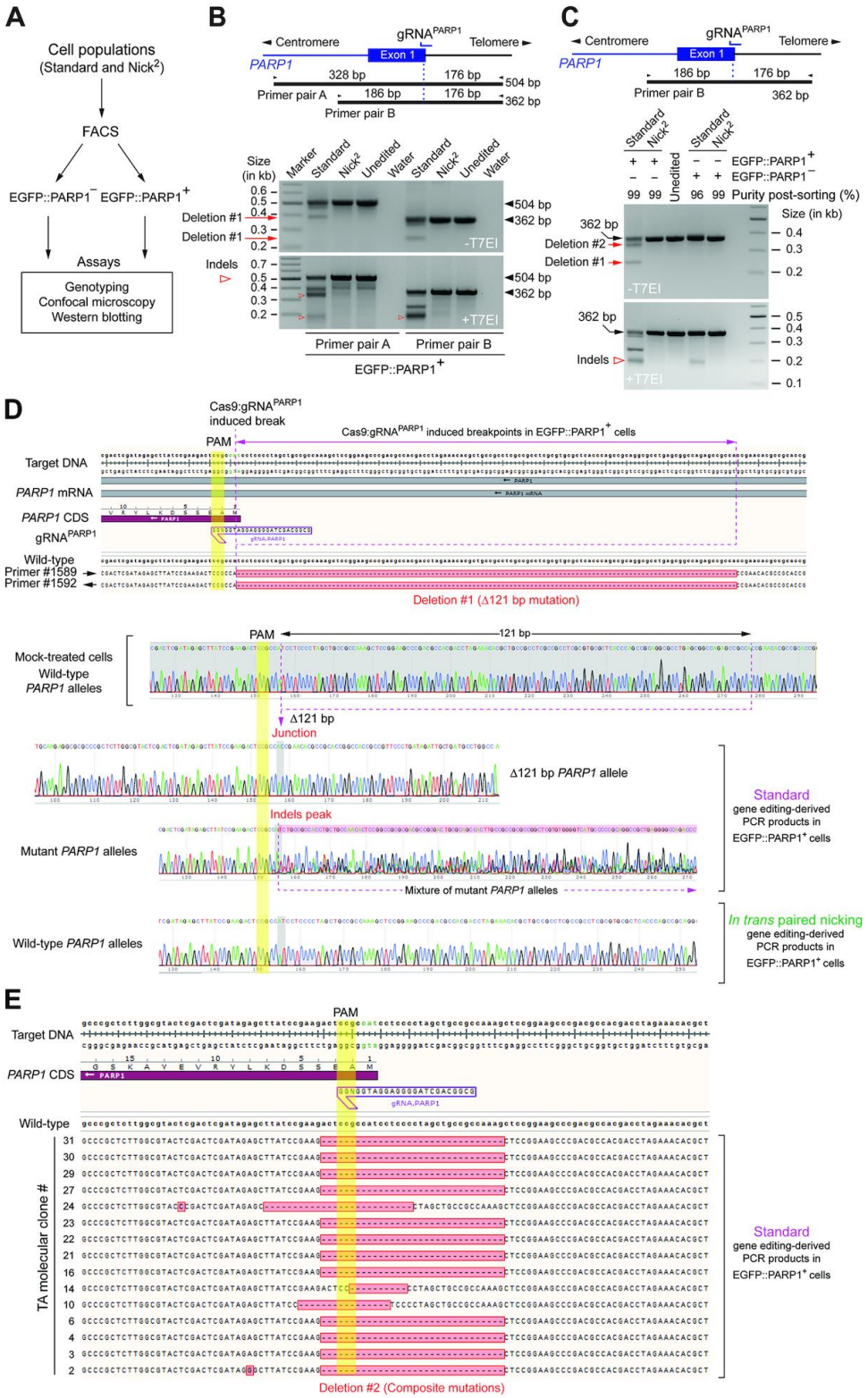


Figure 4. Characterization of *PARP1* alleles in cell populations subjected to standard versus *in trans* paired nicking gene editing. (**A**) Overview of the experimental design. HeLa cell populations subjected to SSB-mediated in trans paired nicking and DSB-mediated standard gene editing were sorted in their respective EGFP::PARP1^{-} and EGFP::PARP1^{+} populations. Each of these cell fractions was next characterized at the DNA and protein levels by the indicated assays. (**B** and **C**) Examination of *PARP1* mutagenesis after gene editing based on DSBs versus SSBs. Untreated and T7EI-treated PCR products spanning the gRNA^{PARP1} target site provided evidence for large deletions and small indels, respectively, in EGFP::PARP1^{+} cells generated by standard gene editing (panel **B**). Indels were equally detected in EGFP::PARP1^{-} cells exposed to standard gene editing (panel **C**). DNA species diagnostic for SpCas9:gRNA^{PARP1}-induced deletions and indels are marked with arrows and open arrowheads, respectively. (**D**) Sequence analysis of the *PARP1* target region in gene edited cells. Top panel, Sanger sequencing of the low molecular weight amplicons shown in panel **B** (-T7EI, primer pair **B**) with forward and reverse primers revealing the presence of a 121-bp deletion at target sequences in EGFP::PARP1^{+} cells that underwent standard gene editing. The *PARP1* proximal deletion

breakpoint coincides with the predicted SpCas9:gRNA^{PARP1} cleaving position. Bottom panel, chromatograms corresponding to *PARP1* alleles in EGFP::PARP1^{+} cells engineered by standard gene editing and in trans paired nicking. Chromatograms corresponding to wild-type *PARP1* and to the 121-bp *PARP1* deletion are also displayed. (**E**) Characterization of additional *PARP1* deletion products. The *PARP1* species with a molecular weight between unmodified and 121 bp-deleted alleles (Deletion #2) presented various mutations as determined by TA cloning and sequence analysis.

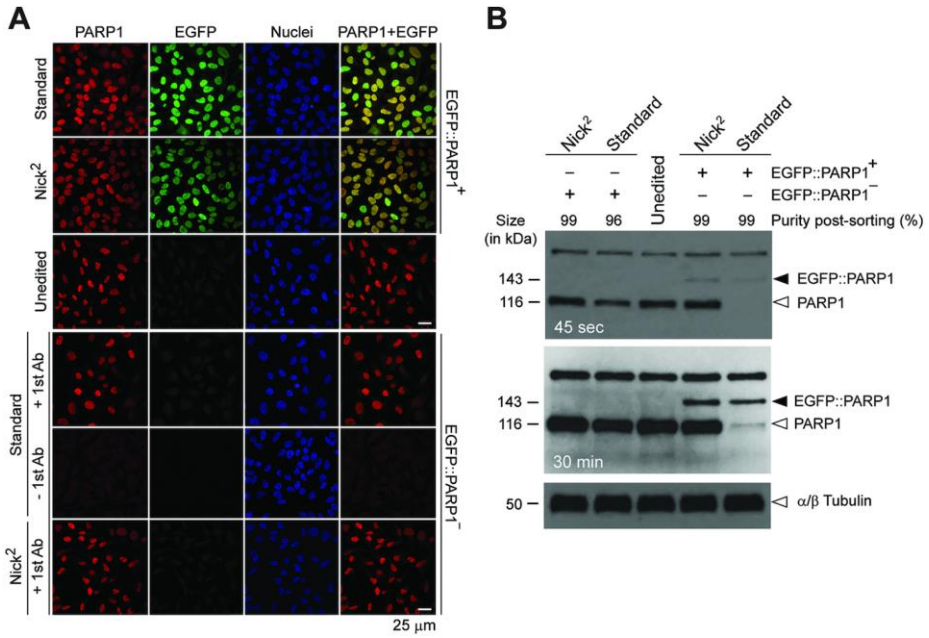


Figure 5. Examination of PARP1 protein status after gene editing triggered by DSBs versus SSBs. (**A**) Confocal microscopy analysis of HeLa cells expressing untagged and EGFP-tagged PARP1. Confocal microscopy of EGFP::PARP1^{+} and EGFP::PARP1^{-} cells confirming co-localization of PARP1 and EGFP in the nuclei of the former cell populations engineered by in trans paired nicking or standard gene editing. Nuclei were counterstained with DAPI. Unedited HeLa cells served as negative controls. Specimens of EGFP::PARP1^{-} cells not incubated with the primary PARP1-specific antibody (-1st Ab) provided for an additional staining control. (**B**) Western blot analysis of HeLa cells expressing untagged and EGFP-tagged PARP1. Western blotting of EGFP::PARP1^{+} and EGFP::PARP1^{-} cells exposing a striking reduction in the amounts of endogenous PARP1 antigen exclusively in EGFP::PARP1^{+} cells generated through standard DSB-dependent gene editing (open arrowhead). Properly sized EGFP::PARP1 fusion products were detected in both EGFP::PARP1^{+} cell populations (solid arrowhead). Unedited HeLa cells served as negative controls. α/β Tubulin antigens served as internal protein loading controls.

In trans paired nicking achieves seamless editing of essential iPSC genomic sequences

The OCT4 transcription factor is essential for human embryogenesis (49) and for the genetic circuitry underpinning pluripotent stem cell states (50,51). For these reasons, it is a coveted gene-editing target. Yet, especially at its termini, *OCT4* shares substantial homology with several of its pseudogenes (**Figure 6A** and **B**). These multiple-copy sequences make the identification of suitable gRNAs hard or impossible (**Figure 6A** and **Supplementary Figure S13**). Hence, we next sought to compare the performance of the different gene editing strategies in a challenging gene-editing model involving tagging *OCT4* at its last exon using gRNAs that lack *OCT4* specificity. To this end, HeLa cells and iPSCs were transfected with conventional pDonor^{OCT4} or target site-modified pDonor$^{OCT4.TS}$, each mixed with plasmids coding for SpCas9:gRNA$^{OCT4.1}$ or SpCas9^{D10A}:gRNA$^{OCT4.1}$ (**Figure 6B**). Colony-formation assays showed that, when compared to single nicking and standard gene editing approaches, in trans paired nicking comprising SSB formation at *OCT4* and donor templates led to higher numbers of puromycin-resistant colonies regardless of the cell type (**Figure 6C**). Similar results were obtained in independent iPSC transfections in which an additional gRNA was included (**Supplementary Figure S14**). Crucially, genomic DNA analysis of randomly isolated iPSC colonies readily revealed that in trans paired nicking achieved a much higher precision in *OCT4* targeting than the DSB-dependent approaches (**Supplementary Figure S15A** and **S15B**). Multicolour FISH-based molecular karyotyping (COBRA-FISH) revealed that neither iPSCs subjected to *in trans* paired nicking nor iPSCs exposed to the DSB-dependent protocols harboured overt chromosomal rearrangements (n = 6; **Figure 7A**). Possibly, this outcome is the result of a strong selection against iPSCs that had initially been exposed to multiple DSBs. Related with this, robust mutagenesis at gRNA$^{OCT4.1}$ target sites located in off-target chromosomal locations (**Figure 7B**) was readily detected in iPSC populations subjected to DSB-dependent gene editing (**Figure 7C**). The

fact that gRNA target sequences in *OCT4* pseudogenes overlap with coding cellular genes, further compounds the genotype of SpCas9:gRNA$^{OCT4.1}$-treated cells (**Figure 7B** and **C**).

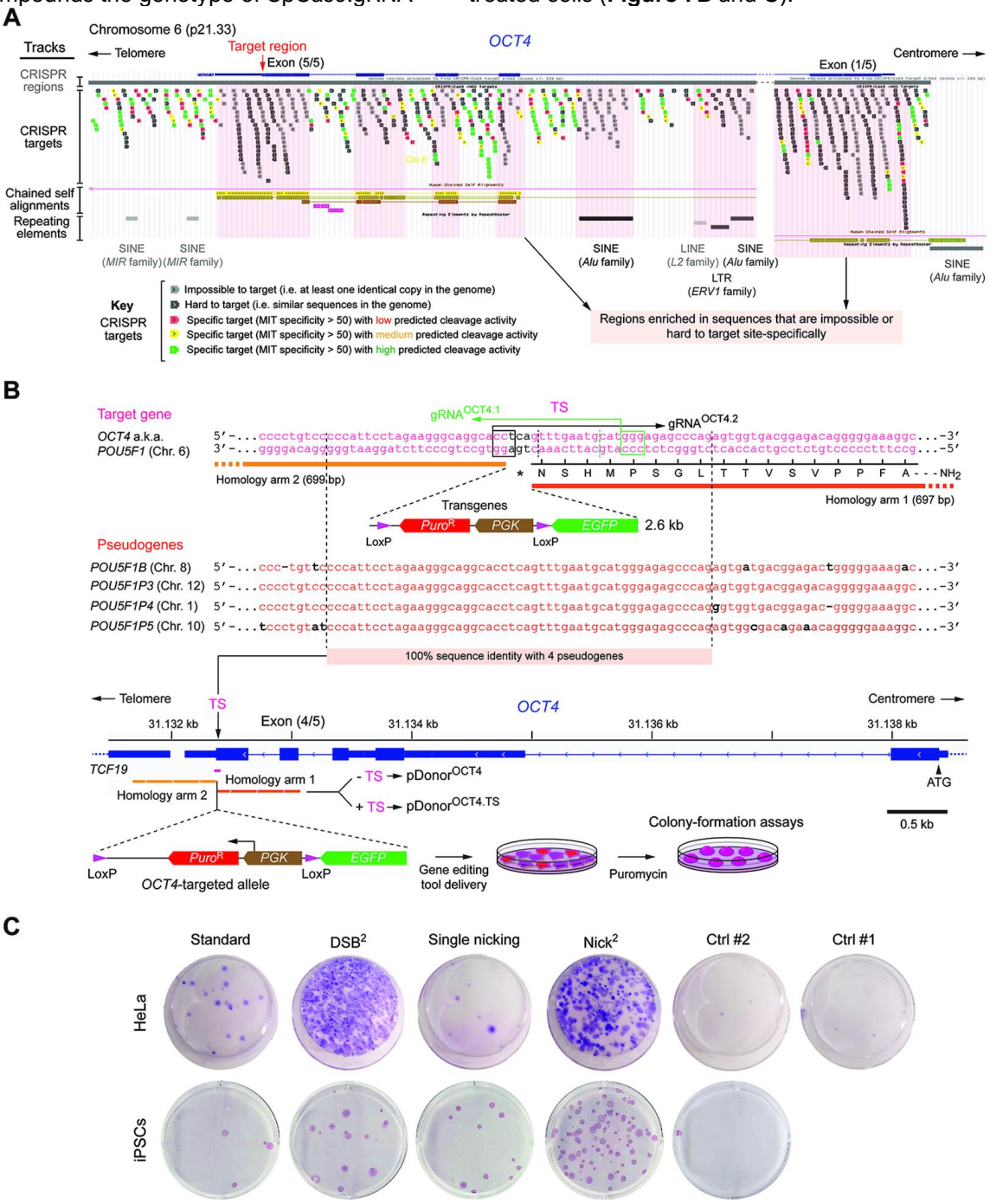


Figure 6. Homology-directed *OCT4* gene editing based on cleaving or nicking CRISPR complexes. (**A**) The *OCT4* genomic region. All potential *S. pyogenes* CRISPR-SpCas9 target sites, as defined by 20-mer spacers and canonical NGG PAMs, are colour-coded according to their predicted target site specificity and activity (CRISPR targets track). Genomic features sharing full or partial sequence identity with *OCT4* are highlighted as duplications and repeats (chained self-alignments and repeating elements tracks, respectively). Tracks annotations were retrieved from the UCSC Genome Browser, Assembly GRCh38/hg38. (**B**) The *OCT4* target region. The *OCT4* terminal nucleotides are drawn in relation to similar sequences present in its pseudogenes and in donor plasmids pDonor^{OCT4} and pDonor$^{OCT4.TS}$. The former and latter constructs lack and contain, respectively, gRNA target sites (TS) flanking the targeting module. The target sites are indicated by horizontal arrows and boxed nucleotides (PAMs). Donor constructs are built to knock-in a floxed positive-selection cassette plus an *EGFP* reporter into *OCT4* loci. The Cre-mediated excision of the selection cassette generates a traceable OCT4::EGFP fusion product exclusively in accurately targeted iPSCs. (**C**) *OCT4* gene editing. Colony-formation assays for detecting stably transfected cells. iPSCs and HeLa cells were co-transfected with conventional pDonor^{OCT4} or target site-modified pDonor$^{OCT4.TS}$ templates each mixed with constructs expressing SpCas9:gRNA$^{OCT4.1}$ or SpCas9^{D10A}:gRNA$^{OCT4.1}$. After puromycin selection, alkaline phosphatase and Giemsa staining identified genetically modified colonies of iPSCs and HeLa cells, respectively.

The generation of DSBs at *OCT4* pseudogenes (**Figure 7C**) raises the possibility for the insertion of *OCT4*-targeting donor DNA at these off-target genomic positions due to the partial homology between them and donor DNA (**Supplementary Figure S15C**). A junction PCR assay devised to investigate this possibility did not detect donor DNA insertions at *OCT4* pseudogenes in puromycin-resistant iPSC clones (n = 22) randomly isolated from cultures subjected to *in trans* paired nicking (**Supplementary Figure S15C** and **D**).

Previous experiments in pluripotent stem cells (i.e. human embryonic stem cells and iPSCs) revealed that *in trans* paired nicking using SpCas9^{D10A}:gAAVS1 complexes yields higher gene targeting frequencies than those achieved by standard gene-editing involving SpCas9:gAAVS1 (12). Similar *AAVS1* gene targeting experiments performed in the iPSC line used in the current study were consistent with these earlier findings (**Supplementary Figure S16**). To investigate whether chromosomal rearrangements were detectable in these iPSCs soon after their exposure to CRISPR complexes, we performed orthogonal HTGTS analysis on cell populations exposed to SaCas9:Sa–gRAG1.1 alone or together with SpCas9:gAAVS1 or SpCas9^{D10A}:gAAVS1 complexes (**Supplementary Figure S17**). The orthogonal HTGTS assay detected translocations exclusively in iPSCs that had been co-treated with SaCas9:Sa-gRAG1.1 and SpCas9:gAAVS1 nucleases (**Supplementary Figures S18** and **S19**). When compared with the orthogonal HTGTS experiments performed in aneuploid HEK293T cells (**Figure 1C**, **Supplementary Figures S3** and **S4**), the overall lower frequencies of translocations detected in iPSCs might have resulted from their diploid character and/or lower exposure to CRISPR complexes (compare **Supplementary Figures S2** with **Supplementary Figure S17**). Crucially, in line with the orthogonal HTGTS experiments in HEK293T cells, this data support that SpCas9^{D10A} nickases trigger less chromosomal rearrangements than their SpCas9 counterparts, in this case, in diploid iPSCs (**Supplementary Figures S18** and **S19**).

To complement the characterization of gene-edited iPSCs (**Figure 7** and **Supplementary Figure S15**), we set-up a quantitative specificity assay in which Cre-mediated OCT4::EGFP assembly reports on precise gene editing in iPSCs (**Figure 8A**). The results from this functional genetic assay confirmed the strikingly different *OCT4* targeting levels achieved by nicking versus cleaving CRISPR complexes. In particular, in contrast to the single nick-dependent and DSB-dependent approaches, induction of SSBs at acceptor and donor DNA results in efficient targeted gene editing in viable iPSCs (**Figure 8B**). Our results suggest that exposing iPSCs to nicking as opposed to cleaving CRISPR complexes overcomes a strong negative selection against *OCT4*-edited iPSCs. These results are in agreement with previous experiments showing that even very few DSBs, including those made by SpCas9 nucleases, can significantly reduce the division and survival rates of PSCs (12,52–54). Finally, dual-colour confocal microscopy and flow cytometry analyses confirmed proper EGFP tagging of the endogenous OCT4 protein in iPSCs subjected to in trans paired nicking, at both the population and clonal levels (**Figure 8C** and **D**, respectively). Importantly, these OCT4::EGFP-expressing iPSCs were equally capable of differentiating along the three embryonic germ layers (**Figure 8E** and **Supplementary Figure S20**).

In conclusion, unwarranted genotypes and deleterious phenotypic traits created by CRISPR-SpCas9 nucleases during gene knock-in procedures are mostly avoided by using *in trans* paired nicking genome editing.

DISCUSSION

There are some concerns regarding the application of genome editing technologies. This is especially so when these applications are directed towards biotechnologies and genetic therapies (55). In part these concerns stem from the fact that, regardless of their specificity, programmable nucleases generate DSBs that are prone to large-scale and small-scale mutagenic events (4–10). In this regard, programmable nuclease-induced DSBs are particularly problematic, hence avoided, at multiple-copy sequences and/or at sequences needed for proper cell functioning or overall viability. As corollary, DSB-dependent genome editing substantially limits the editable genome. Moreover, in mammalian diploid cells, nuclease-induced homologous chromosome rearrangements (10) and allelic mutations potentiate cell transformation events and gene-dose unbalances, respectively. Equally insidious are the recent findings that DSB-induced nonsense mutations can trigger transcriptional compensatory mechanisms that further confound genotype-phenotype associations (56–58).

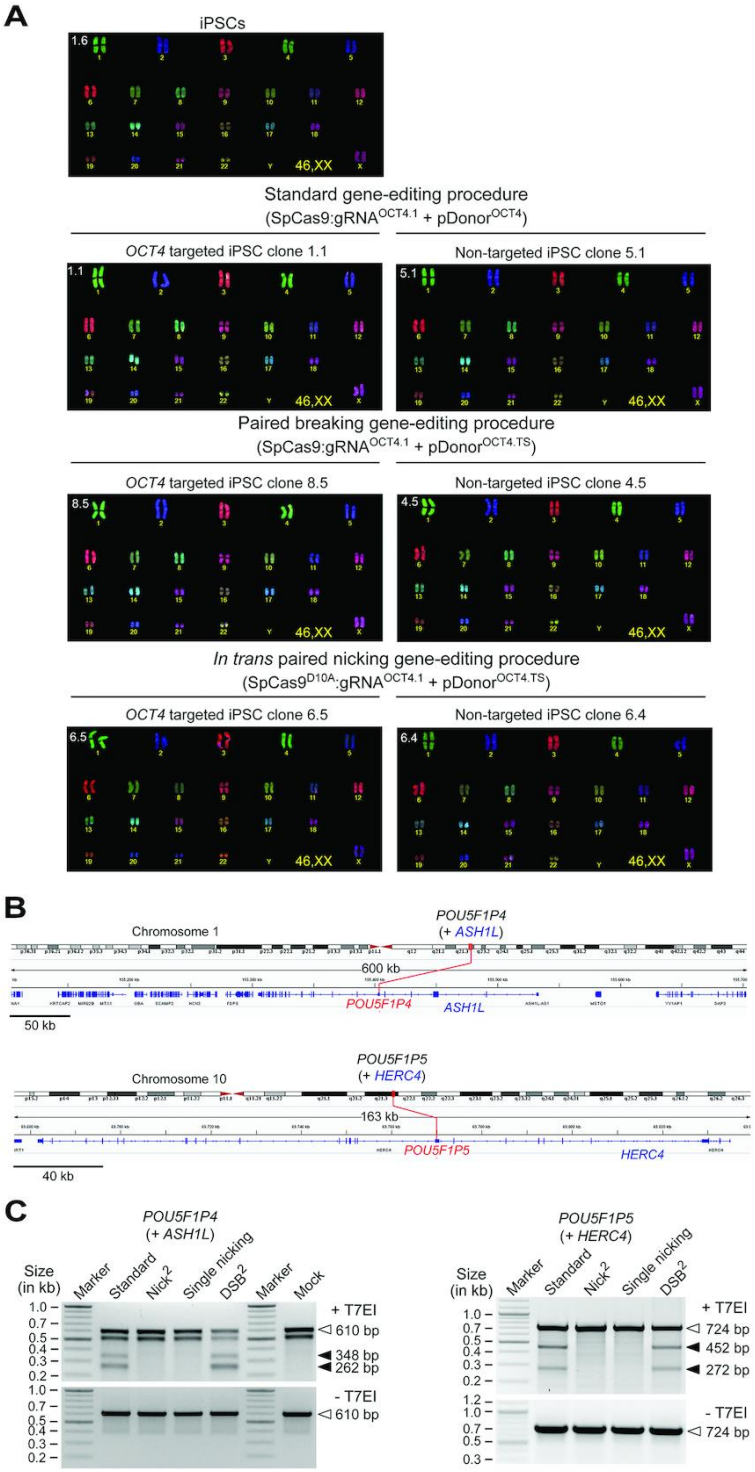


Figure 7. Characterization of iPSCs after *OCT4* gene editing using cleaving versus nicking CRISPR complexes. (**A**) Karyotyping of genetically modified iPSC clones. Overview of COBRA-FISH analysis of parental iPSCs and individual targeted and non-targeted clones showing a seemingly normal diploid karyotype (46,XX). Each clone was isolated after adding puromycin to iPSC populations subjected to the indicated gene editing strategies. (**B**) Chromosomal and genomic coordinates of *POU5F1P4* and *POU5F1P5*. The former and latter *OCT4* pseudogenes overlap with nucleotide sequences from *ASH1L* (ASH1-like histone lysine methyltransferase) and *HERC4* (HECT and RLD domain containing E3 ubiquitin protein ligase 4), respectively. *ASH1L* codes for a member of the trithorax group of transcriptional activators and is ubiquitously expressed in over 25 tissues; *HERC4* belongs to the HERC family of ubiquitin ligases and is ubiquitously expressed in over 25 tissues. As a result, indels generated at *OCT4* pseudogenes inevitably create additional genotypic complexity in target cell populations whose, cell type-specific, phenotypic consequences are difficult to predict and assess. (**C**) Comparing genome-disrupting events at *OCT4* gRNA target sites located at off-target chromosomal positions. T7EI-based genotyping assays were performed on DNA from puromycin-resistant iPSC populations expanded after *OCT4*-targeting experiments involving the indicated gene editing procedures. T7EI-specific products diagnostic for mutant alleles generated by NHEJ-mediated DSB repair are pinpointed by closed arrowheads;

products corresponding to intact alleles are instead indicated by open arrowheads in untreated and T7EI-treated samples. Marker, GeneRuler DNA Ladder Mix molecular weight marker.

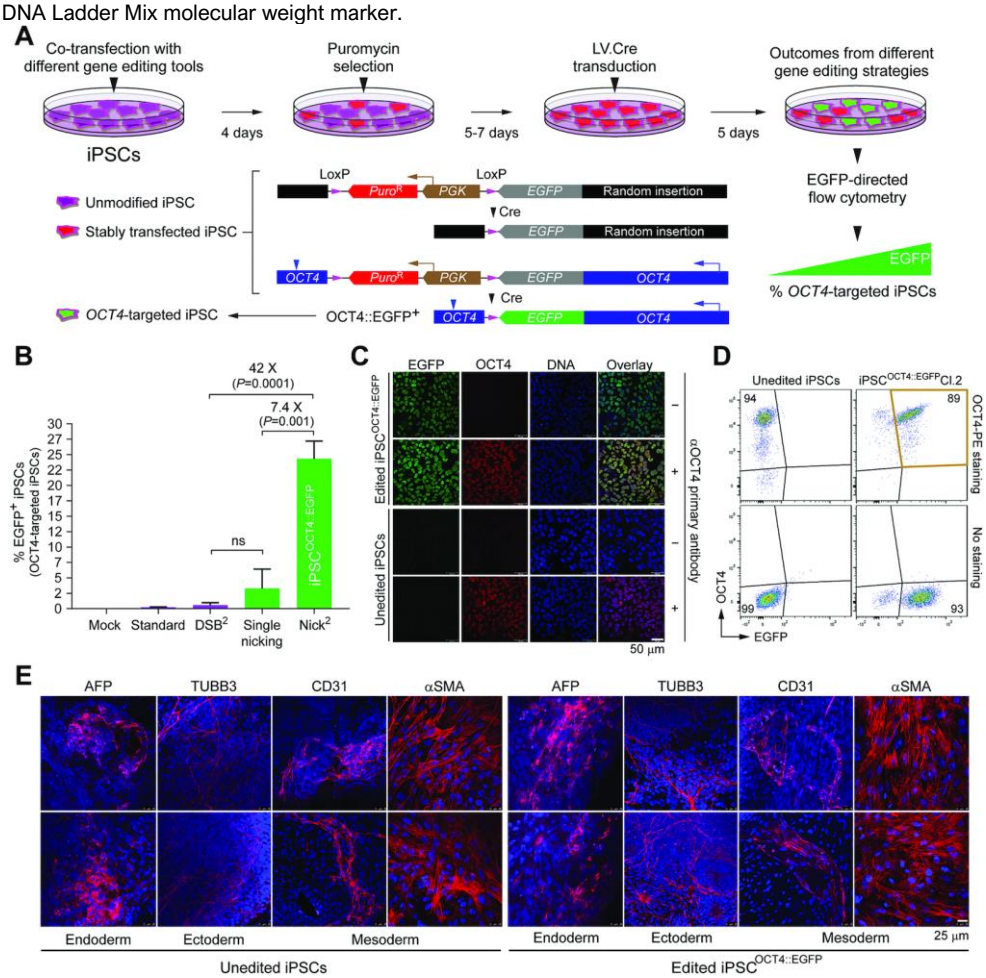


Figure 8. Comparing the accuracy of OCT4 gene editing after delivering cleaving versus nicking CRISPR complexes. (**A**) Genetic assay for determining *OCT4* targeting frequencies. iPSCs co-transfected with plasmid combinations corresponding to each of the four different gene editing strategies, were sequentially exposed to puromycin and Cre recombinase. *OCT4*-targeted iPSCs expressing Cre-derived OCT4::EGFP fusion products report accurate genome-modifying events. The Cre recombinase was delivered by transducing iPSCs with lentiviral vector LV.Cre at a multiplicity-of-infection of 10 physical particles per cell. (**B**) Comparing the performance of *OCT4* gene editing strategies in iPSCs. The frequencies of *OCT4*-targeted iPSCs expressing OCT4::EGFP were determined by EGFP-directed flow cytometry. Data are shown as mean ± S.D. of independent biological replicates. Significance was calculated with two-tailed Student's *t* tests (n = 3); ns, non-significant. (**C**) Confocal microscopy analysis of *OCT4* edited iPSCs. OCT4::EGFP-expressing iPSCs engineered through in trans paired nicking and Cre delivery (iPSC$^{OCT4::EGFP}$) were subjected to indirect and direct fluorescence microscopies for detecting OCT4 and EGFP, respectively. Nuclei were stained with DAPI. Nuclear localization of OCT4::EGFP is highlighted by the merging of the three fluorescence signals. Unedited iPSCs served as negative controls. iPSC and iPSC$^{OCT4::EGFP}$ populations that were not incubated with the OCT4-specific primary antibody served as staining controls. (**D**) Flow cytometric analysis of *OCT4* edited iPSCs. Flow cytometry of iPSC clone 2 isolated from an iPSC$^{OCT4::EGFP}$ population confirming OCT4 and EGFP co-labelling (coloured quadrant). Unedited iPSCs served as controls. Cultures of parental iPSCs and iPSC$^{OCT4::EGFP}$ clone 2 that were not exposed to the PE-conjugated OCT4 antibody were used as staining controls. (**E**) Testing multi-lineage differentiation capacity of iPSC populations expressing OCT4::EGFP. Immunofluorescence microscopy analysis of iPSC$^{OCT4::EGFP}$ cells differentiated into cellular lineages representative of endoderm, ectoderm and mesoderm. Unedited iPSCs served as differentiation controls. Markers for each germ layer are indicated. Nuclei were stained with DAPI.

We report that concomitant SSB formation at target and donor DNA by CRISPR-SpCas9 nickases elicits accurate and non-disruptive gene editing, including at loci associated with haploinsufficiency and essentiality. This DSB-free *in trans* paired nicking approach prevented the loss of gene-edited cells due to the disruption of a functional protein motif or a pluripotency supporting gene in iPSCs. The observed difficulty in isolating iPSCs edited at *OCT4* after CRISPR-SpCas9 delivery is in line with the essentiality of this gene in safeguarding stem cell phenotypes (49–51) and with earlier experiments showing that gene targeting frequencies at *OCT4* are very low. Indeed, gene editing of iPSCs using TALENs and the herein used pDonor^{OCT4} construct, did not yield any correctly targeted clone (0/48) (28). In another study, gene editing of human embryonic stem cells deploying SpCas9 and donor templates containing the same 'homology arms' of pDonor^{OCT4}, resulted in only 8 correctly targeted clones (8/288) (59). In

contrast to these studies, viable and correctly targeted iPSC clones were readily isolated after targeting *OCT4* with pDonor$^{OCT4.TS}$ and SpCas9^{D10A} (21/22) (**Supplementary Figure S15B**). Importantly, in trans paired nicking gene editing introduces a low mutagenic load into target cell populations by minimizing NHEJ-mediated chromosomal disruption of allelic and non-allelic target sequences, such as those in *OCT4* and its pseudogenes, respectively. These multiple-copy gRNA target sites, are likely to have exacerbated the difficulty in isolating *OCT4*-targeted iPSCs after SpCas9 delivery (**Figure 8B** and **Supplementary Figure S15B**) as pluripotent stem cells are particularly prone to DSB-induced cell cycle arrest and apoptosis (12,52–54). There are other experimental data linking detrimental genome editing outcomes to target sequences associated with copy number variations. In particular, genome-wide CRISPR-SpCas9 library screens have demonstrated that DSBs mapping in amplified genomic regions create false-positive hits of gene essentiality in cancer cell lines (60,61).

Notwithstanding the fact that nicking CRISPR complexes are significantly less mutagenic than their cleaving counterparts at both target and off-target sites, they can nonetheless trigger DNA disruptions if, for example, an advancing replication fork collapses after hitting the SSB product (42). In the present work, by using orthogonal HTGTS assays, we have provided experimental evidence for such events in mammalian cells (**Figure 1C** and **Supplementary Figures S3** and **S4**). These events should be most problematic at off-target sites. In this regard, it will be worth investigating whether *in trans* paired nicking is amenable to RNA-guided nickases built on high-specificity Cas9 scaffolds (62).

Although the *OCT4* edited iPSC clones analysed lacked donor DNA insertions at SSB-susceptible *OCT4* pseudogenes (**Supplementary Figure S15D**), unwanted knock-ins at genomic regions exhibiting high homology with donor DNA constitute a possible limitation of *in trans* paired nicking. Therefore, whenever possible, this risk should be minimized by avoiding SSB formation at such potential off-target regions and/or reducing the extent of homology between them and donor DNA (63). Conversely, assuring SSB formation at donor DNA and multiple-copy homologous sequences might offer the prospect for co-editing these recurrent regions in the genome without attendant large-scale chromosomal mutations and rearrangements.

In conclusion, HDR-mediated gene editing through *in trans* paired nicking offers high specificity and low mutagenicity, which, as a result, mostly preserves cellular genotypes and phenotypes. Moreover, the coordinated nicking of donor and acceptor HDR templates boosts the versatility of CRISPR-based gene editing by substantially enlarging the fraction of candidate gRNAs that can become operational, regardless of their *a priori* specificity profiles. The seamless and scarless character of *in trans* paired nicking should be particularly beneficial in instances in which precise and predictable genetic interventions are crucial. Examples include modelling or rescuing disease traits in stem cells (64) and functionally dissecting genomic sequences by multiplexed knock-in of donor DNA libraries (65). Finally, *in trans* paired nicking might expand the 'editable genome' to different types of repetitive elements shedding light on this large and variegated portion of the functionally unknown genomic 'dark matter' (66).

DATA AVAILABILITY

All data generated and analysed in this study are included in the article and its supplementary files. Additional raw datasets are available from the corresponding author on reasonable request. The Gene Expression Omnibus (GEO) datasets corresponding to the orthogonal HTGTS analyses are available via accession code GSE135064.

SUPPLEMENTARY DATA

Supplementary Data are available online at doi: 10.1093/nar/gkz1121.

ACKNOWLEDGEMENTS

The authors thank Martijn Rabelink, Steve Cramer and Hans Vrolijk (Department of Cell and Chemical Biology, LUMC, Leiden, the Netherlands) for performing p24^{gag} ELISA titrations of lentiviral vector particles, assistance with the research logistics and making available the ColorProc source code, respectively. The authors also thank Elisa Giacomelli and Dorien Ward-van Oostwaard for sharing their iPSC culture protocols and Christian Freund (Department of Anatomy and Embryology, LUMC, Leiden, the Netherlands) for making available the Nucleofector 2b-device.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 765269; and from the Dutch Prinses Beatrix Spierfonds (W.OR16-13); X.C. and Q.W. receive support from Ph.D. fellowships from the China Scholarship Council-Leiden University Joint Scholarship Programme.
Conflict of interest statement. None declared.

REFERENCES

1. Chandrasegaran,S. and Carroll,D. (2016) Origins of programmable nucleases for genome engineering. *J. Mol. Biol*., **428**, 963-989.
2. Maggio, I. and Gonçalves,M.A. (2015) Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends Biotechnol*., **33**, 280-291.
3. Chang,H.H.Y., Pannunzio,N.R., Adachi,N. and Lieber,M.R. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell. Biol*., **18**, 495-506.
4. Kosicki,M., Tomberg,K. and Bradley,A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol*. 2018, **36**, 765-771.
5. Cradick,T.J., Fine, E.J., Antico,C.J. and Bao,G. (2013) CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*., **41**, 9584-9592.
6. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol*., **31**, 822-826.
7. Cho,S.W., Kim,S., Kim,Y., Kweon,J., Kim,H.S., Bae,S. and Kim,J.S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*., **24**, 132-141.
8. Lin,Y., Cradick,T.J., Brown,M.T., Deshmukh,H., Ranjan,P., Sarode,N., Wile,B.M., Vertino,P.M., Stewart,F.J. and Bao,G. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res*., **42**, 7473-7485.
9. Kuscu,C., Arslan,S., Singh,R., Thorpe,J. and Adli,M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol*., **32**, 677-683.
10. Frock,R.L., Hu,J., Meyers,R.M., Ho,Y.J., Kii,E. and Alt,F.W. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol*., **33**, 179-186.
11. Holkers,M., Maggio,I., Henriques,S.F., Janssen,J.M., Cathomen,T. and Gonçalves MA. (2014) Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat. Methods*., **11**, 1051-1057.
12. Chen,X., Janssen,J.M., Liu,J., Maggio,I., 't Jong,A.E., Mikkers,H.M., Gonçalves,M.A. (2017) In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun*., **8**, 657.
13. Zhang,J.P., Li,X.L., Li,G.H., Chen,W., Arakaki,C., Botimer,G.D., Baylink,D., Zhang,L., Wen,W., Fu,Y.W. *et al*. (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol*., **18**, 35.
14. McConnell Smith,A., Takeuchi,R., Pellenz,S., Davis,L., Maizels,N., Monnat,R.J.Jr. and Stoddard,B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. USA*., **106**, 5099-5104.
15. Metzger,M.J., McConnell-Smith,A., Stoddard,B.L. and Miller,A.D. (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res*., **39**, 926-935.
16. Ramirez,C.L., Certo,M.T., Mussolino,C., Goodwin,M.J., Cradick,T.J., McCaffrey,A.P., Cathomen,T., Scharenberg,A.M. and Joung,J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res*., **40**, 5560-5568.
17. Wang,J., Friedman,G., Doyon,Y., Wang,N.S., Li,C.J., Miller,J.C., Hua,K.L., Yan,J.J., Babiarz,J.E., Gregory,P.D. *et al*. (2012) Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res*., **22**, 1316-1326.

18. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
19. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al*. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.
20. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. (2013) and Church,G.M. RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
21. Doudna,J.A. and Charpentier,E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.
22. Mali,P., Aach,J., Stranges,P.B., Esvelt,K.M., Moosburner,M., Kosuri,S., Yang,L. and Church,G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol*., **31**, 833-838.
23. 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. *et al*. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380-1389.
24. van Nierop,G.P., de Vries,A.A., Holkers,M., Vrijsen,K.R. and Gonçalves,M.A. (2009) Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Res*., **37**, 5725-5736.
25. Nakajima,K., Zhou,Y., Tomita,A., Hirade,Y., Gurumurthy,C.B. and Nakada,S. (2018) Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res*., **28**, 223-30.
26. Paulsen B.S., Mandal P.K., Frock R.L., Boyraz B., Yadav R, Upadhyayula S., Gutierrez-Martinez P., Ebina W., Fasth A., Kirchhausen T. *et al*. (2017) Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing. *Nat. Biomed. Eng.*, **1**, 878–888.
27. Zhang,M., D'Aniello,C., Verkerk,A.O., Wrobel,E., Frank,S., Ward-van Oostwaard,D., Piccini,I., Freund,C., Rao,J., Seebohm,G. *et al*. (2014) Recessive cardiac phenotypes in induced pluripotent stem cell models of Jervell and Lange-Nielsen syndrome: disease mechanisms and pharmacological rescue. *Proc. Natl. Acad. Sci. USA*., **111**, E5383-E5392.
28. Hockemeyer,D., Wang,H., Kiani,S., Lai,C.S., Gao,Q., Cassady,J.P., Cost,G.J., Zhang,L., Santiago,Y., Miller,J.C. *et al*. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol*., **29**, 731-734.
29. Chen,X., Liu,J., Janssen,J.M. and Gonçalves,M.A. (2017) The chromatin structure differentially impacts high-specificity CRISPR-Cas9 nuclease strategies. *Mol. Ther. Nucleic Acids*, **8**, 558-563.
30. Kleinstiver,B.P., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Topkar,V.V., Zheng,Z. and Joung,J.K. (2015) Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol*., **33**, 1293-1298.
31. Hu,J., Meyers,R.M., Dong,J., Panchakshari,R.A., Alt,F.W. and Frock,R.L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc*., **11**, 853-871.
32. Brielmeier,M., Béchet,J.M., Falk,M.H., Pawlita,M., Polack,A. and Bornkamm,G.W. (1998) Improving stable transfection efficiency: antioxidants dramatically improve the outgrowth of clones under dominant marker selection. *Nucleic Acids Res*., **26**, 2082-2085.
33. Szuhai,K. and Tanke,H.J. (2006) COBRA: combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. *Nat. Protoc*., **1**, 264-275.
34. Hsiau,,T., Maures,T., Waite,K., Yang,J., Kelso,R., Holden,K. and Stoner,R. (2018) Inference of CRISPR Edits from Sanger Trace Data. biorXiv, 10.1101/251082v2, 14-January-2018, pre-print: not peer-reviewed.
35. Bae,S., Park,J. and Kim,J.S. (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, **30**, 1473-1475.
36. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R. *et al*. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol*., **34**, 184-191.
37. Schwartz,S., Kent,W.J., Smit,A., Zhang,Z., Baertsch,R., Hardison,R.C., Haussler,D. and Miller,W. (2003) Human-mouse alignments with BLASTZ. *Genome Res*., **13**, 103-107.

38. Pelascini,L.P., Janssen,J.M. and Gonçalves,M.A. (2013) Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. *Hum Gene Ther*., **24**, 78-96.
39. Pelascini,L.P. snf Gonçalves,M.A. (2014) Lentiviral vectors encoding zinc-finger nucleases specific for the model target locus HPRT1. *Methods Mol. Biol*., **1114**, 181-199.
40. Gabriel,R., von Kalle,C. and Schmidt,M. (2015) Mapping the precision of genome editing. *Nat. Biotechnol*. , **33**, 150-152.
41. Wienert,B., Wyman,S.K., Richardson,C.D., Yeh,C.D., Akcakaya,P., Porritt,M.J.., Morlock,M., Vu,J.T., Kazane,K.R., Watry,H.L. *et al*. (2019) Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* , **364**, 286-289.
42. Kuzminov,A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. USA* , **98**, 8241-8246.
43. Casper,J., Zweig,A.S., Villarreal,C., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Karolchik,D. *et al*. (2018) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res*., **46**, D762-D769.
44. Foster,E.R. and Downs,J.A. (2005) Histone H2A phosphorylation in DNA double-strand break repair. *FEBS J*. , **272**, 3231-3240.
45. Holliday,R. (1964) A mechanism for gene conversion in fungi. *Genet. Res*. **5**:283–304.
46. Celeste,A., Difilippantonio,S., Difilippantonio,M.J., Fernandez-Capetillo,O., Pilch,D.R., Sedelnikova,O.A., Eckhaus,M., Ried,T., Bonner,W.M. and Nussenzweig,A. (2003) H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. *Cell*., **114**, 371-383.
47. Hanzlikova,H., Gittens,W., Krejcikova,K., Zeng,Z. and Caldecott,K.W. (2017) Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res*., **45**, 2546-2557.
48. Ronson,G.E., Piberger,A.L., Higgs,M.R., Olsen,A.L., Stewart,G.S., McHugh,P.J., Petermann,E. and Lakin ND. (2018) PARP1 and PARP2 stabilise replication forks at base excision repair intermediates through Fbh1-dependent Rad51 regulation. *Nat. Commun*. , **9**, 746.
49. Fogarty,N.M.E., McCarthy,A., Snijders,K.E., Powell,B.E., Kubikova,N., Blakeley,P., Lea,R., Elder,K., Wamaitha,S.E., Kim,D. *et al*. (2017) Genome editing reveals a role for OCT4 in human embryogenesis. *Nature*, **550**, 67-73.
50. Takahashi,K., Tanabe,K., Ohnuki,M., Narita,M., Ichisaka,T., Tomoda,K. and Yamanaka S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861-872.
51. Yilmaz,A., Peretz,M., Aharony,A., Sagi,I. and Benvenisty,N. (2018) Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol*., **20**, 610-619.
52. Liu,J.C., Lerou,P.H. and Lahav,G. (2014) Stem cells: balancing resistance and sensitivity to DNA damage. *Trends Cell. Biol*., **24**, 268–274.
53. Ihry,R.J., Worringer,K.A., Salick,M.R., Frias,E., Ho,D., Theriault,K., Kommineni,S., Chen,J., Sondey,M., Ye,C. et al. (2018) p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med*., **24**, 939-946.
54. Haapaniemi,E., Botla,S., Persson,J., Schmierer,B. and Taipale,J. (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med*., **24**, 927-930.
55. Smalley,E. (2018) As CRISPR-Cas adoption soars, summit calls for genome editing oversight. *Nat. Biotechnol*., **36**, 4852018.
56. El-Brolosy,M.A., Kontarakis,Z., Rossi,A., Kuenne,C., Günther,S., Fukuda,N., Kikhi,K., Boezio,G.L.M., Takacs,C.M., Lai,S.L. *et al*. (2019) Genetic compensation triggered by mutant mRNA degradation. *Nature*, **568**, 193-197.
57. Ma,Z., Zhu,P., Shi,H., Guo,L., Zhang,Q., Chen,Y., Chen,S., Zhang,Z., Peng,J. and Chen,J. (2019) PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature*, **568**, 259-263.
58. Tuladhar,R., Yeu,Y., Piazza,J.T, Tan,Z., Clemenceau,J.R., Wu,X., Barrett,Q., Herbert,J., Mathews,D.H., Kim,J. *et al*. (2019) CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat. Commun*., **10**, 4056.
59. Zhu,Z., Verma,N., González,F., Shi,Z.D. and Huangfu,D. (2015) A CRISPR/Cas-mediated selection-free knockin strategy in human embryonic stem cells. *Stem Cell Rep*., **4**, 1103-1111.

60. Munoz,D.M., Cassiani,P.J., Li,L., Billy,E., Korn,J.M., Jones,M.D., Golji,J., Ruddy,D.A., Yu,K., McAllister,G. *et al*. (2016) CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov*., **6**, 900-913.
61. Gonçalves,E., Behan,F.M., Louzada,S., Arnol,D., Stronach,E.A., Yang,F., Yusa,K., Stegle,O., Iorio,F. and Garnett,M.J. (2019) Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss of fitness effects. *Genome Biol*., **20**, 27.
62. Chen,X. and Gonçalves,M.A. (2018) DNA, RNA, and Protein Tools for Editing the Genetic Information in Human Cells. *iScience*, **6**, 247-263.
63. Deng, C. and Capecchi,M.R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell Biol*., **12**, 3365-3371.
64. Bellin,M., Marchetto,M.C., Gage,F.H. and Mummery,C.L.(2012) Induced pluripotent stem cells: the new patient? *Nat. Rev. Mol. Cell. Biol*., **13**, 713-726.
65. Findlay,G.M., Boyle,E.A., Hause,R.J., Klein,J.C. and Shendure,J. (2014) Saturation editing of genomic regions by multiplex homology-directed repair. *Nature*, **513**, 120-123.
66. Sedlazeck,F.J., Lee,H., Darby,C.A. and Schatz,M.C. (2018) Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet*., **19**, 329-346.

Chapter 3

Precise and broad scope genome editing based on high-specificity Cas9 nickases

Qian Wang[1], Jin Liu[1], Josephine M. Janssen[1], Marie Le Bouteiller[2], Richard L. Frock[2] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands;
[2]Department of Radiation Oncology, Division of Radiation and Cancer Biology, Stanford University School of Medicine, 269 Campus Dr. Stanford, CA 94305, USA

Nucleic Acids Research, 2021, Vol.49, No.2, ***1173–1198***
doi:10.1093/nar/gkaa1236

ABSTRACT

RNA-guided nucleases (RGNs) based on CRISPR systems permit installing short and large edits within eukaryotic genomes. However, precise genome editing is often hindered due to nuclease off-target activities and the multiple-copy character of the vast majority of chromosomal sequences. Dual nicking RGNs and high-specificity RGNs both exhibit low off-target activities. Here, we report that high-specificity Cas9 nucleases are convertible into nicking Cas9^{D10A} variants whose precision is superior to that of the commonly used Cas9^{D10A} nickase. Dual nicking RGNs based on a selected group of these Cas9^{D10A} variants can yield gene knockouts and gene knock-ins at frequencies similar to or higher than those achieved by their conventional counterparts. Moreover, high-specificity dual nicking RGNs are capable of distinguishing highly similar sequences by 'tiptoeing' over pre-existing single base-pair polymorphisms. Finally, high-specificity RNA-guided nicking complexes generally preserve genomic integrity, as demonstrated by unbiased genome-wide high-throughput sequencing assays. Thus, in addition to substantially enlarging the Cas9 nickase toolkit, we demonstrate the feasibility in expanding the range and precision of DNA knockout and knock-in procedures. The herein introduced tools and multi-tier high-specificity genome editing strategies might be particularly beneficial whenever predictability and/or safety of genetic manipulations are paramount.

INTRODUCTION

RNA-guided nucleases (RGNs) based on prokaryotic CRISPR–Cas9 adaptive immune systems consist of ribonucleoprotein complexes made of single guide RNAs (gRNAs) and Cas9 nucleases (1). RGNs are programmable nucleases in that they can be tailored to cleave specific DNA sequences whose recognition involves sequential protein–DNA and RNA–DNA interactions. Firstly, the Cas9 component binds to a so-called protospacer adjacent motif (PAM) on the DNA (2). The PAM of the prototypic *Streptococcus pyogenes* Cas9 (SpCas9) nuclease and that of its orthologue *Staphylococcus aureus Cas9* (SaCas9) nuclease, reads NGG and NNGRRT, respectively (3,4). Secondly, hybridization of the 5' end of the gRNA (spacer) to a normally 20 nucleotide-long sequence (protospacer) located next to the PAM ultimately triggers double-stranded DNA break (DSB) formation through the allosteric activation of the two Cas9 nuclease domains, i.e. RuvC-like and HNH (1). Hence, RGNs bypass the need for protein engineering owing to their RNA-based programmability and, as such, constitute versatile and powerful tools for changing specific nucleotide sequences amidst large eukaryotic genomes (1,5). Commonly, such genome editing maneuvers yield gene knockouts and, in the presence of exogenous donor DNA, gene knock-ins resulting from non-homologous end joining (NHEJ) and homology-directed repair (HDR) of site-specific DSBs, respectively (1,5).

Despite the far-reaching appeal of RGN technologies, major concerns regarding their use are, however, off-target DNA cleavage and associated collateral effects, e.g. chromosomal sequence disruptions and translocations (6–12). Off-target activities result from the fact that, often, RGNs remain cleaving-proficient even when several mismatches exist between gRNA and genomic sequence(s). This is especially so if the mismatches locate distally to the PAM (7–9). Moreover, although to a lesser degree than NGG, certain non-canonical PAMs (e.g. NAG) can also be engaged by *S. pyogenes* Cas9 and lead to off-target DSB formation when located next to sequences fully or partially complementary to the gRNA spacer (7–9,12–14).

RGN off-target activities have prompted an increasing number of Cas9 mutagenesis screens based on rational design and directed evolution principles whose results include an expanding portfolio of Cas9 variants with enhanced target site specificities (15). A parallel, broadly applicable, approach for reducing off-target activities involves using nicking RGN (nRGN) pairs containing sequence- and strand-specific Cas9 nucleases (nickases) generated by disabling either the RuvC-like (Cas9^{D10A}) or the HNH (Cas9^{H840A}) domains (3,16,17). The simultaneous induction of single-stranded DNA breaks (SSBs) at offset positions in opposite target DNA chain by pairs of these nicking RGNs (dual nRGNs) yields a targeted DSB (18,19). Crucially, SSBs made at off-target sites by individual dual nRGN pair members are mostly repaired through conservative, non-mutagenic, DNA repair processes (20,21). Notably, when compared to regular RGNs containing Cas9, dual nRGNs harboring Cas9^{D10A} offer a higher target-site selection density and, hence, wider genomic space coverage. This follows from the fact that the effective spacing separating the bipartite target sites of dual nRGNs is relatively broad (up to ~100 bp) widening the range for locating suitable PAMs (18,19). Moreover, dual nRGNs containing Cas9^{D10A} can sometimes induce higher target DNA cleaving activities when compared to their corresponding monomeric RGNs (22).

Presumably, this results from the fact that such dual nRNGs bypass the need for a functional RuvC-like domain, which of the two SpCas9 nuclease domains, seems to be the least catalytically active in mammalian cells (22).

In this study, we start by investigating whether a representative panel of SpCas9 nucleases with enhanced specificities, i.e. SpCas9-KA (23), SpCas9-KARA (23), eSpCas9(1.1) (23), Sniper-Cas9 (24), SpCas9-HF1 (25), evoCas9 (26) and xCas9-3.7 (27), are convertible into functional nicking forms. In these experiments, the activities and specificities of the respective nRGNs were compared with those containing the conventional Cas9^{D10A} nickase. Subsequently, we asked whether these new enzymes are operational as dual nRGNs for triggering gene knockouts and gene knock-ins in human cells, including induced pluripotent stem cells (iPSCs). We report that high-specificity SpCas9 proteins vary greatly in their permissiveness to the incorporation of the RuvC-disabling D10A mutation. Indeed, the phosphodiester bond cleaving efficiencies achieved by these RNA-programmable nickases, in their single and dual nRGN formats, varies from lower to higher than those obtained via their respective, unmodified, Cas9^{D10A}-containing counterparts. Importantly, the identified high-activity Cas9^{D10A} nickases endow single and dual nRGNs with specificities that are superior to those conferred by the unmodified Cas9^{D10A} protein.

MATERIAL AND METHODS

Cells

Human cervix carcinoma (HeLa) cells and human embryonic kidney 293T (HEK293T) cells (both from American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific; Cat. No.: 41966029) supplemented with 5% and 10% fetal bovine serum ultra-low endotoxin (FBS; Biowest; Cat. No.: S1860500), respectively. The generation and characterization of H2AX::mCherry^{+}, TURQ2 and H27 cells were described elsewhere (14,28,29). All these reporter HeLa cell-derived cell lines were maintained in DMEM containing 5% FBS. The human iPSCs used in this study (LUMC0020iCTRL06) were generated and characterized elsewhere (28). The iPSCs were maintained in feeder-free Essential 8 Medium (E8; Thermo Fisher Scientific; Cat. No.: A1517001) supplemented with 25 U ml^{-1} penicillin and 25 μg ml^{-1} of streptomycin (Thermo Fisher Scientific; Cat. No.: 15140122). The cells were passaged as small clumps using 0.5 mM ethylenediaminetetraacetic acid (EDTA) (Invitrogen; Cat. No: 15575020) diluted 1:1000 in Dulbecco's phosphate-buffered saline (DPBS; Thermo Fisher Scientific; Cat. No.: 14190094) every three to four days and were re-plated in wells of six-well plates (Greiner Bio-One; Cat. No.: 662160) containing E8 medium supplemented with a 1:200 dilution of RevitaCell (Thermo Fisher Scientific; Cat. No.: A2644501). All the cell culture vessels used for iPSCs culture in this work were coated with Vitronectin Recombinant Human Protein (VTN-N; Thermo Fisher Scientific; Cat. No.: A14700) diluted 1:100 to a final concentration of 5 ng ml^{-1} in DPBS for at least 1 h at room temperature (RT). The various cell types were kept at 37°C in a humidified-air 10% CO2 atmosphere except for iPSCs, which were instead maintained in a humidified-air 5% CO2 atmosphere. The cells used in this work were tested for the absence of mycoplasma.

Recombinant DNA

The isogenic expression plasmids containing the open reading frames of the SpCas9 nucleases and SpCas9 nickases under the control of the same hybrid CAG promoter and rabbit *β-globin* polyadenylation signal, were assembled on the basis of the indicated previously published constructs and BB36_pCAG.Cas9eSp(1.1)-D10A.bGHpA, AL65_pEX-A128.partialCas9-eSp(1.1).K1003-R1060, AL66_pEX-A128.partialCas9-eSp(1.1).K1003, BA59_pUC57.start-Cas9-HF1-D10A, AL68_pEX-A258.Cas9-evo(partial) and BA16_pU.CAG.dSaCas9.rBGpA. The codes and names of the expression plasmids encoding SpCas9 nucleases and nickases generated in this study are gathered in **Supplementary Table S1**. The annotated maps and nucleotide sequences of BB36_pCAG.Cas9eSp(1.1)-D10A.bGHpA, AL65_pEX-A128.partialCas9-eSp(1.1).K1003-R1060, AL66_pEX-A128.partialCas9-eSp(1.1).K1003, BA59_pUC57.start-Cas9-HF1-D10A, AL68_pEX-A258.Cas9-evo(partial) and BA16_pU.CAG.dSaCas9.rBGpA are available in pages 1–14 of the Supplementary Information. The amino acid sequences of nickases encoded by AB65_pU.CAG.Cas9-D10A.rBGpA (14), AP76_pU.CAG.Cas9-D10A-K848A.rBGpA, AP70_pU.CAG.Cas9-D10A-K848A-R1060A.rBGpA, AA69_pU.CAG.Cas9-eSp(1.1)-D10A.rBGpA.2NLS, AE70_pU.CAG.SniperCas9-D10A.rBGpA, BB37_pU.CAG.Cas9-HF1-D10A.rBGpA, AP74_pU.CAG.Cas9-evo-D10A.rBGpA and AT85_pU.CAG.xCas9-3.7-D10A.rBGpA are depicted in pages 15–22 of the Supplementary Information.

Constructs AW01_pU.CAG.Cas9-eSp(1.1).rBGpA (30) and BB36_pCAG.Cas9eSp(1.1)-D10A.bGHpA were digested with BshTI and Eco32I. Subsequently, the 7378-bp backbone fragment from AW01_pU.CAG.Cas9-eSp(1.1).rBGpA (30) and the 1982-bp insert fragment from BB36_pCAG.Cas9eSp(1.1)-D10A.bGHpA were extracted from agarose gel and ligated together, leading to the generation of construct AA69_pU.CAG.Cas9-eSp(1.1)-D10A.rBGpA.2NLS encoding eSpCas9(1.1)^{D10A}. Next, AW01_pU.CAG.Cas9-eSp(1.1).rBGpA (30) and AA69_pU.CAG.Cas9-eSp(1.1)-D10A.rBGpA.2NLS were digested with Eco72I and BsmI, after which, the 8509-bp backbone fragments were isolated from agarose gel and dephosphorylated with FastAP (Thermo Fisher Scientific; Cat. No.: EF0651) for 1 h at 37°C according to the specifications of the manufacturer. The 851-bp insert fragments encoding SpCas9-KA and SpCas9-KARA were extracted from agarose gel after digesting AL65_pEX-A128.partialCas9-eSp(1.1).K1003-R1060 and AL66_pEX-A128.partialCas9-eSp(1.1).K1003 with Eco72I and BsmI. Subsequently, the resulting insert fragments were ligated to the dephosphorylated vector backbone from AW01_pU.CAG.Cas9-eSp(1.1).rBGpA (30) or that from AA69_pU.CAG.Cas9-eSp(1.1)-D10A.rBGpA.2NLS. These maneuvers led to the assembly of expression constructs AP75_pU.CAG.Cas9-K848A.rBGpA, AP76_pU.CAG.Cas9-D10A-K848A.rBGpA, AP69_pU.CAG.Cas9-K848A-R1060A.rBGpA and AP70_pU.CAG.Cas9-D10A-K848A-R1060A.rBGpA encoding SpCas9-KA, SpCas9-KA^{D10A}, SpCas9-KARA and SpCas9-KARA^{D10A}, respectively. To generate expression plasmids encoding Sniper-Cas9 and Sniper-Cas9^{D10A}, constructs AV62_pU.CAG.Cas9.rBGpA (30) and AB65_pU.CAG.Cas9-D10A.rBGpA (14) were digested with SdaI and Eco72I. The resulting 6673-bp backbone fragments were then extracted from agarose gel and dephosphorylated as above-indicated. Next, plasmid AV72_pCMV.Sniper-Cas9.bGHpA (Addgene plasmid #113912) was digested with SdaI and Eco72I, after which, the 2542-bp insert fragment was ligated to the dephosphorylated vector backbones from AV62_pU.CAG.Cas9.rBGpA (30) and AB65_pU.CAG.Cas9-D10A.rBGpA (14), yielding constructs AE69_pU.CAG.SniperCas9.rBGpA and AE70_pU.CAG.SniperCas9-D10A.rBGpA, respectively. For generating the construct encoding SpCas9-HF1^{D10A}, plasmids AV64_pU.CAG.Cas9-HF1.rBGpA (30) and BA59_pUC57.start-Cas9-HF1-D10A were digested with SacI and BstZ17I. Subsequently, the 9039-bp backbone fragment from AV64_pU.CAG.Cas9-HF1.rBGpA (30) and the 261-bp insert fragment from BA59_pUC57.start-Cas9-HF1-D10A were isolated from agarose gel and ligated together, leading to the expression construct BB37_pU.CAG.Cas9-HF1-D10A.rBGpA. To assemble expression plasmids encoding evoCas9 and evoCas9^{D10A}, constructs AV62_pU.CAG.Cas9.rBGpA (30) and AB65_pU.CAG.Cas9-D10A.rBGpA (14) were digested with SalI and BamHI and, after agarose gel extraction, the 7750-bp backbone fragments were dephosphorylated. Next, construct AL68_pEX-A258.Cas9-evo(partial) was digested with SalI and BamHI, after which, the 1465-bp insert fragment was isolated from agarose gel and ligated to the dephosphorylated vector backbones from AV62_pU.CAG.Cas9.rBGpA (30) and AB65_pU.CAG.Cas9-D10A.rBGpA (14), resulting in constructs AP73_pU.CAG.Cas9-evo.rBGpA and AP74_pU.CAG.Cas9-evo-D10A.rBGpA, respectively. To generate expression plasmids encoding xCas9-3.6, xCas9-3.6^{D10A}, xCas9-3.7 and xCas9-3.7^{D10A}, AV62_pU.CAG.Cas9.rBGpA (30) and AB65_pU.CAG.Cas9-D10A.rBGpA (14) were digested with SdaI and BshTI and the 5309-bp backbone fragments were then extracted from agarose gel and dephosphorylated. In parallel, AE65_pCMV.xCas9-3.6.HSV-TKpA (Addgene plasmid #108384) and AE66_pCMV.xCas9-3.7.HSV-TKpA (Addgene plasmid #108379) were digested with SdaI and BshTI and the 3908-bp insert fragments were then isolated from agarose gel and ligated to the dephosphorylated vector backbone from AV62_pU.CAG.Cas9.rBGpA (30) or that from AB65_pU.CAG.Cas9-D10A.rBGpA (14). These manoeuvres led to the assembly of AT82_pU.CAG.xCas9-3.6.rBGpA, AT83_pU.CAG.xCas9-3.6-D10A.rBGpA, AT84_pU.CAG.xCas9-3.7.rBGpA, and AT85_pU.CAG.xCas9-3.7-D10A.rBGpA encoding xCas9-3.6, xCas9-3.6^{D10A}, xCas9-3.7 and xCas9-3.7^{D10A}, respectively. The generation of the construct expressing nicking SaCas9D10A was carried out as follows. Plasmids BA15_pCAG.SaCas9.rBGpA (31) and BA16_pU.CAG.dSaCas9.rBGpA were digested with BcuI and Kpn2I, after which, the 5063-bp backbone and 3316-bp insert fragments, respectively, were isolated from agarose gel and ligated to each other yielding BA31_pU.CAG.SaCas9-D10A.rBGpA. The expression plasmids coding for gRNAs used in this work were assembled by inserting annealed oligonucleotide pairs indicated in **Supplementary Table S2** into BvelI-digested AY56_pUCBM21.U6.opt-sgRNA.BvelI-stuffer (32). AV85_pSa-gRAG1.1 (14) and AM51_pUCBM21.U6.gRNAI-SceI.1 (30), encoding *RAG1*-specific Sa-gRNA1.1 and an irrelevant, non-targeting gRNA, respectively, have been described previously (14,30).

Cell transfections

With the exception of iPSCs, all other cell types were seeded in the cell culture vessels indicated in **Supplementary Tables S3–S26**. At ~16–24 h after seeding, the cells were transfected with the aid of 1 mg ml^{-1} 25 kDa linear polyethyleneimine (PEI, Polysciences) solution (pH 7.4). The cell numbers, the amounts of PEI, DNA (in ng) and 150 mM NaCl (in µl) as well as the compositions of each DNA mixture corresponding to the different transfection reactions are specified in **Supplementary Tables S3–S26**. Prior to transfection the plasmids were first diluted in 150 mM NaCl (Merck), after which, the appropriate amount of the PEI solution was added to each of the transfection reactions. After vigorously vortexing for about 10 s, the transfection mixtures were incubated for 15 min at RT to let PEI–DNA complexes form. The resulting transfection mixtures were then directly added into the culture media of the target cells and, after 6 h, the transfection media were substituted by regular culture media. The transfections of iPSCs were done by using Lipofectamine Stem Transfection Reagent (Thermo Fisher Scientific, Cat. No.: STEM00003) according to the manufacturer's protocols. In brief, cells were seeded in wells of 24-well plates coated with Vitronectin with the culture media refreshed at least 2 h prior to transfection. The cell numbers, the amounts of Lipofectamine Stem Transfection Reagent (in µl), DNA (in ng) as well as the compositions of each of the DNA mixtures corresponding to the different transfection reactions are specified in **Supplementary Table S27**. The plasmid mixtures and the appropriate amounts of Lipofectamine Stem Transfection Reagent were diluted in 25 µl of Opti-MEM medium (Gibco; Cat. No.: 31985-047) in 1.5-ml sterile Eppendorf tubes. After mixing, by gently pipetting, the resulting transfection reactions were incubated at RT for 10 min and were then directly added into the culture media of the target iPSCs. The transfection media were replaced with regular iPSC culture medium 24 h post-transfection.

Flow cytometry

Gene knockout frequencies in transfected cell populations were determined by flow cytometry of reporter-negative cells at 10 days post-transfection and, with the exception of the experiments presented in **Figure 1** and **Supplementary Figure S1**, were normalized for initial transfection efficiencies on a per sample basis by reporter-directed flow cytometry at 3 days post-transfection. The flow cytometry analyses were carried out by using a BD LSR II flow cytometer (BD Biosciences). In brief, cells were trypsinized, washed with PBS and resuspended in PBS supplemented with 0.5% bovine serum albumin (BSA) and 2 mM EDTA (pH 8.0). Parental non-transfected cells were used as negative controls to set background fluorescence. At least 10 000 viable single cells were acquired per sample. Data were analyzed with the aid of FlowJo 10.5.0 software (Tree Star).

Western blotting

Cells were lysed with Laemmli buffer consisting of 8.0% glycerol, 3% sodium dodecyl sulfate (SDS) and 200 mM Tris–HCl (pH 6.8), followed by boiling at 100°C for 5 min. Protein concentrations were measured by a DC™ protein assay kit (Bio-Rad; Cat. No.: 5000111) according to the manufacturer's instructions. Equal amounts of proteins were loaded and separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Afterwards, the resolved proteins were transferred onto 45-µm polyvinylidene difluoride (PVDF) membrane (Merck Millipore; Cat. No.: IPVH00010). Next, 5% non-fat dry milk dissolved in Tris-buffered saline with 0.1% Tween 20 (TBST) was used to block the membrane at RT for 1 h. Membranes were incubated overnight at 4°C with the respective primary antibodies recognizing S. pyogenes Cas9 (Abcam; Cat. No.: ab191468), α/β Tubulin (Cell Signaling; Cat. No.: 2148), and GAPDH (Merck Millipore; Cat. No.: MAB374) diluted 1:1000 in TBST supplemented with 5% BSA. Subsequently, the membranes were washed with TBST thrice and probed with secondary antibodies specific for mouse IgG (Sigma-Aldrich; Cat. No.: NA931V) or rabbit IgG (Cell Signaling; Cat. No.: 7074S) diluted 1:5000 in TBST containing 1% non-fat dry milk at RT for 2 h. Clarity™ Western ECL Substrate (Bio-Rad; Cat. No.: 1705060) was applied for signal detection using the ChemiDoc Imaging System (Bio-Rad; Cat. No.: 17001402).

Testing gene-editing tools at alternate chromatin states

Cultures of HEK.EGFP$^{TetO.KRAB}$ cells (30), were either not treated or treated with doxycycline (Dox) at a final concentration of 200 ng ml^{-1} starting 7 days prior to transfection (**Supplementary Table S19**). After a sub-culture period of 10 days, HEK.EGFP$^{TetO.KRAB}$ cells that were kept in the presence or absence of Dox (200 ng ml^{-1}), were incubated for an additional 7-day period, after which, the frequencies of EGFP-negative cells were determined by flow cytometry.

Target-site genotyping assays
Genotyping assays based on the mismatch-sensing T7 endonuclease I (T7EI), were performed for the assessment of NHEJ-derived indel formation at target sequences. In brief, genomic DNA was extracted by using the DNeasy Blood & Tissue Kit (Qiagen; Cat. No.: 69506) according to the manufacturer's instructions. Next, the various target sites were amplified with the aid of the primers listed in **Supplementary Tables S28** and **S29**. The cycling conditions and PCR mixture compositions used are specified in **Supplementary Tables S28** and **S30–S33**. The resulting amplicons were subjected to cycles of denaturation and reannealing to form heteroduplexes using the thermocycling parameters indicated in **Supplementary Table S34**. Subsequently, 10 μl of reannealed samples were treated with 0.5 μl (5U) of T7EI (New England Biolabs; Cat. No.: M0302) at 37°C for 15 min and were analysed by agarose gel electrophoresis. Parallel samples of reannealed amplicons not treated with T7EI served as negative controls. After electrophoresis, untreated and T7EI-treated amplicons were detected by using the Gel-Doc XR+ system and the ImageLab 4.1 software (both from Bio-Rad).

Clonal analysis for assessing gene knock-ins at *OCT4* in HeLa cells
HeLa cells were transfected as indicated in **Supplementary Table S24**. At 3 days post-transfection, the cells were transferred into wells of six-well plates (Greiner Bio-One) and were subsequently exposed to 1 μg ml^{-1} puromycin (Invitrogen, Cat. No.: A11138-03) for 7 days. The resulting puromycin-resistant HeLa clones were identified through colony-formation assays using standard Giemsa or Crystal violet staining protocols. In addition, parallel cultures of puromycin-resistant HeLa cell populations were seeded at a density of 0.3 cells per well in wells of 96-well plates (Greiner Bio-One). The resulting single cell-derived clones were then sub-cultured for ~3 weeks in DMEM supplemented with 5% FBS, 1 μg ml^{-1} puromycin, 50 nM α-thioglycerol (Sigma-Aldrich; Cat. No.: M6145) and 0.02 nM bathocuproinedisulfonic acid disodium salt (Sigma-Aldrich; Cat. No.: B1125). Subsequently, genomic DNA of randomly collected single cell-derived clones was extracted and analysed by junction PCR using Phire™ Tissue Direct PCR Master Mix (Thermo Fisher Scientific, Cat. No.: F-107L) according to the manufacturer's protocols. The PCR primer pairs, composition of the PCR mixtures and cycling parameters are specified in **Supplementary Tables S35** and **S36**, respectively.

Quantification of *OCT4* gene targeting frequencies in iPSCs
The transfection of iPSCs was carried out as indicated under 'Cell transfections' and in **Supplementary Table S27**. At 2 days post-transfection, the iPSCs were transferred into new wells of 24-well plates (Greiner Bio-One) and were subsequently expanded into wells of six-well plates (Greiner Bio-One) for 5–7 days in the presence of 0.5 μg ml^{-1} puromycin in E8 Medium containing 25 U ml^{-1} penicillin and 25 μg ml^{-1} streptomycin. The resulting puromycin-resistant iPSC colonies were identified through colony-formation assays using the leukocyte AP kit and protocol (Sigma-Aldrich; Cat. No.: 86R-1KT). In addition, parallel cultures of puromycin-resistant iPSC populations were further expanded for quantification of *OCT4* gene targeting frequencies. In brief, puromycin-resistant iPSC populations resulting from the different *OCT4* gene targeting strategies were reseeded in wells of 24-well plates at a density of 40,000 cells per well. The next day, a lentiviral vector expressing the bacteriophage P1 Cre recombinase (LV.Cre) (14) was added to the target iPSCs at a multiplicity-of-infection (MOI) of 20 viral particles per cell. After a 5-day sub-culture period, the frequency of iPSCs expressing OCT4::EGFP, assembled via Cre-mediated recombination, was measured by flow cytometry.

Confocal immunofluorescence microscopy
Cells were fixed in 4% paraformaldehyde (PFA) and were permeabilized in 0.5% Triton X-100 in tris-buffered saline (TBS) pH 7.6 (50 mM Tris–HCl pH 7.6; 150 mM NaCl) at RT for 10 min, after three washes with 0.1% Triton X-100 in TBS (TBST). A blocking solution consisting of TBS, 0.1% Triton X-100, 2% BSA and 0.1% sodium azide was applied to block non-specific antibody binding for 1 h at RT. Next, the cells were incubated with the primary antibodies indicated in **Supplementary Table S37**, diluted in blocking solution for 1 h at RT. The specimens were subsequently subjected to three washes with TBST and the target antigens were probed with fluorochrome-conjugated secondary antibodies diluted in blocking solution for 1 h in the dark at RT (**Supplementary Table S37**). Finally, ProLong Gold Antifade Mounting reagent containing DAPI (Thermo Fisher Scientific; Cat. No.: P36931) was used for mounting samples after three washes with TBST. The fluorescence images were captured with the aid of an upright Leica SP8 confocal microscope (Leica Microsystems) equipped with Leica hybrid detectors, HyD (Leica Microsystems) and were analyzed using LAS X software.

Spontaneous differentiation of iPSCs

OCT4::EGFP^{+} iPSC populations were dissociated into large cell clumps by scrapping after incubating them in PBS/EDTA for 1 min at 37°C. The cell clumps were then cultured in suspension at 37°C for 24 h on low-attachment plates containing culture media E8. Next, the cell clumps were seeded on glass coverslips coated with Vitronectin in culture media supplemented with Revitacell. After 2 days in culture, the medium was changed to differentiation medium DMEM/F12 (Gibco; Cat. No. 31331-028) containing 20% FBS. The differentiation medium was replenished every 2–3 days during the following 3 weeks. Immunofluorescence staining was carried out to detect the markers for mesoderm, ectoderm and endoderm (**Supplementary Table S37**). The targeted markers for these embryonic germ layers were, α-smooth muscle actin (α-SMA), tubulin β3 class III (TUBB3) and α-fetoprotein (AFP), respectively.

Preparation of genomic DNA for orthogonal HTGTS analysis

The isolation of genomic DNA used for orthogonal HTGTS analysis was detailed elsewhere (14). In brief, HEK293T cells transfected as indicated in **Supplementary Table S26**, were collected at 36 h post-transfection and were resuspended in freshly prepared lysis buffer containing 200 mM NaCl, 10 mM Tris–HCl (pH 7.4), 2 mM EDTA (pH 8.0), 0.2% SDS and 200 ng ml−1 proteinase K (Thermo Fisher Scientific; Cat. No.: #EO0491). After overnight incubation at 56°C, genomic DNA was precipitated by adding isopropanol to a final concentration of 50% and then washed with 1 ml of 70% ethanol. After centrifugation at 13 000 × g for 5 min at 4°C, genomic DNA pellets were dissolved in TE buffer (10 mM Tris–HCl pH 8.0 and 1 mM EDTA pH 8.0) for at least 2 h at 56°C. The assessment of bait and prey chromosomal DNA breaks at RAG1 and VEGFA alleles in the transfected HEK293T cell populations was done using T7EI-based genotyping assays. To this end, the RAG1 and VEGFA target regions were PCR-amplified with KOD Hot Start DNA Polymerase (Merck Millipore; Cat. No.: 71086-3) and GoTaq G2 Flexi DNA Polymerase (Promega; Cat. No.: M7805) by using the PCR mixtures indicated in **Supplementary Tables S32** and **S33**, respectively. The PCR primer pairs and cycling parameters are specified in **Supplementary Tables S29** and **S31**, respectively. Subsequently, the amplicons were subjected to T7EI treatments for the detection of indels at *RAG1* and *VEGFA* loci.

Assessing genome-wide off-target effects through orthogonal HTGTS analysis

The orthogonal HTGTS analyses on genomic DNA samples extracted from transfected HEK293T cells were performed in a blind fashion. The reagents and protocols used in HTGTS, including the orthogonal HTGTS assay, have been detailed elsewhere (12,14,33). In this work, however, prey/bait sequence alignments were performed against human genome assembly hg38 instead of hg19. In brief, 25-μg genomic DNA samples were sheared in a Bioruptor (Diagenode) with a circulating temperature of 4°C using a low-power setting, i.e. 2 × 30 s pulses intercalated by a cooldown period of 60 s. The biotinylated RAG1A/B-F1 primer (12) was used for LAM-PCR (33). Prior to the ligation of bridge adapters (12,33), the LAM-PCR ssDNA products were purified using streptavidin-coated magnetic beads (ThermoFisher Scientific; Cat. No.: 65002). Barcoded RAG1A/B-F2 I5 and AP2 I7 primers (12) and primers P5–I5 and P7–I7 primers (33) were applied for the nested PCR and final PCR, respectively. The PCR products ranging in size from 500 bp to 1 kb were subsequently purified after agarose gel electrophoresis (Qiagen; Cat. No.: 28706). The Phusion polymerase (ThermoFisher Scientific; Cat. No.: F530L) was used for the synthesis of the various amplicons with the blocking enzyme step being omitted. The HTGTS deep sequencing libraries were run on a Bioanalyzer (Agilent 2100) prior to 250-bp paired end MiSeq sequencing (Illumina; Cat. No.: MS-102-2003). The resulting pooled sequence reads were demultiplexed and trimmed using the selected molecular barcodes and adapter sequences. Finally, each read library was subjected to (i) bait/prey sequence alignments to the human genome assembly hg38, (ii) filtering and (iii) post-pipeline analysis as specified elsewhere (33). Enriched sites are off-target sites found significant in at least one of the total libraries; hotspots are defined as enriched sites found significant in at least 2 out of 3 normalized libraries for each CRISPR complex. Significantly enriched translocation sites and hotspots in sequence read libraries were called using MACS2 (q-value cutoff -10^{-10}), as previously detailed (12).

Target site genotyping by amplicon deep sequencing.

H27 reporter cells and HEK293T cells were exposed to dual nRGNs containing SpCas9^{D10A} or SpCas9^{D10A} variants as indicated under 'Cell transfections' and in **Supplementary Tables S8** and **S9**. At 2 days post-transfection, genomic DNA extracted via the DNeasy Blood & Tissue Kit protocol (Qiagen;

Cat. No.: 69506), was subjected to Illumina MiSeq next generation sequencing for obtaining 100 000 paired end reads from *EGFP* and *H2AX* target sequences in H27 and HEK293T cells, respectively. The NGS procedure was as follows. *EGFP*- and *H2AX*-specific PCR products (254 and 291 bp, respectively), were amplified with Phusion High-Fidelity Polymerase (Thermo Fisher Scientific; Cat. No.: #F-530L) and the PCR mixtures indicated in **Supplementary Table S38**. The primer pairs with adapter tag overhangs and the cycling parameters applied are specified in **Supplementary Tables S39** and **S40**, respectively. After purification using AMPure XP beads (Beckman Coulter; Cat. No.: A63881), the resulting amplicons were subjected to PCR barcoding using Illumina tag-specific primer pairs with unique sequence combinations for demultiplexing and sample identification (**Supplementary Table S41**). The PCR mixtures and cycling parameters used for the preparation of barcoded amplicons are indicated in **Supplementary Tables S42** and **S40**, respectively. After purification using AMPure XP beads, the concentrations of barcoded amplicons were determined by using the Qubit dsDNA HS assay kit (Invitrogen; Cat. No.: Q32854) and a Qubit2.0 fluorometer (Invitrogen). Sample quality control was done by capillarity electrophoresis through a 2100 Bioanalyzer system (Agilent). Finally, libraries of pooled barcoded amplicons were subjected to Illumina MiSeq deep sequencing with the reads corresponding to each individual sample being subsequently analysed with the aid of CRISPResso2 (34). In brief, after demultiplexing, adapter trimming of the paired end MiSeq raw reads (R1 and R2 fastq files) was performed with Cutadapt 2.10. Finally, the alignment of amplicon sequences to reference sequences was carried out by using CRISPResso2 set in the standard NHEJ mode. The codes applied in the CRISPResso2 analysis are available as **Supplementary Information**.

Statistical analyses

With the exception of the genomic DNA samples used in the orthogonal HTGTS analyses, the researchers were not blinded to sample allocation. Data derived from a minimum of three biological replicates were analysed by GraphPad Prism 8.0.1 software package. Statistical significances were analyzed using the tests indicated in the figure legends. P values lower than 0.05 were considered to be statistically significant.

RESULTS

Comparing the performances of standard and high-specificity nucleases

We started by comparing the performance of wild-type SpCas9 with those of SpCas9 mutant variants SpCas9-KA (23), SpCas9-KARA (23), eSpCas9(1.1) (23), Sniper-Cas9 (24), SpCas9-HF1 (25), evoCas9 (26) and xCas9-3.7 (27) (Figure 1A). To this end, TURQ2 reporter cells were transfected with isogenic constructs expressing each of these nucleases (**Figure 1A**) mixed with plasmids synthesizing four different *mTurquoise2*-specific gRNAs. TURQ2 cells (28) contain a constitutively active *mTurquoise2* transgene (35) inserted at the human *AAVS1* 'safe harbor' locus (**Figure 1B**). Hence, mTurquoise2 knockouts, resulting from small insertions and deletions (indels) generated after NHEJ-mediated DSB repair processes, report nuclease activity. To simultaneously confirm the higher specificity of SpCas9 variants over that of SpCas9, an *EGFP*-specific gRNA presenting three mismatches to an *mTurquoise2* sequence (gEGFP.3), was taken along (**Figure 1C**).

Flow cytometric quantification of mTurquoise2-negative cells showed that Sniper-Cas9 was the most consistent nuclease variant in that it yielded the most similar DNA cleaving activities when coupled to each of the four *mTurquoise2*-targeting gRNAs tested. However, once combined with gEGFP.3, Sniper-Cas9 led to off-target activities above background levels (**Figure 1C** and **D**). As expected, the native SpCas9 protein was the least specific enzyme of the panel (**Figure 1C** and **D**). The sub-set formed by the single, double and triple mutants SpCas9-KA, SpCas9-KARA and eSpCas9(1.1), respectively, yielded robust DNA cleaving activities except when combined with gTURQ.2 (**Figure 1C**). Moreover, eSpCas9(1.1) was also significantly less active than SpCas9 when coupled to gTURQ.3 (**Figure 1C**). Contrasting with gTURQ.1, that has a canonical 20-mer spacer fully complementary to the target DNA, the least performing gTURQ.2, similarly to gTURQ.3 and gTURQ.4, has a 21-mer spacer whose 5' terminal guanine does not hybridize to the target sequence. Of notice, such non-canonical gRNAs are common gene-editing reagents due to a strong preference exhibited by frequently used RNA polymerase III promoters for guanines as first transcript nucleotide. Additional experiments performed in EGFP-expressing H27 reporter cells (29) showed that when compared with parental SpCas9, excluding Sniper-Cas9, all other high-specificity SpCas9 nucleases yielded substantially reduced gene knockout levels once coupled to gEGFP.21 whose 21-mer spacer is fully complementary to the target

DNA (**Supplementary Figure S1**). Consistent with our results, gRNAs with 5' non-hybridizing guanines and/or extended spacers were shown to significantly inhibit high-specificity SpCas9 nucleases, including eSpCas9(1.1), SpCas9-HF1 and evoCas9 but less so Sniper-Cas9 (24,26,36–38). Taken together, these data generally confirm the differential performance of the various SpCas9 variants vis-à-vis the wild-type SpCas9 protein in terms of their specificities and compatibilities with different gRNA moieties. Regarding the latter aspect, our data revealed that Sniper-Cas9 is the most compatible with a 5' non-hybridizing guanine whilst evoCas9 the least. Furthermore, our results uncovered an inverse correlation between the increasing number of mutations in the nuclease set formed by SpCas9-KA, SpCas9-KARA and eSpCas9(1.1), and gene knockout frequencies when using gRNAs with 21-mer spacers (**Figure 1C** and **Supplementary Figure S1**).

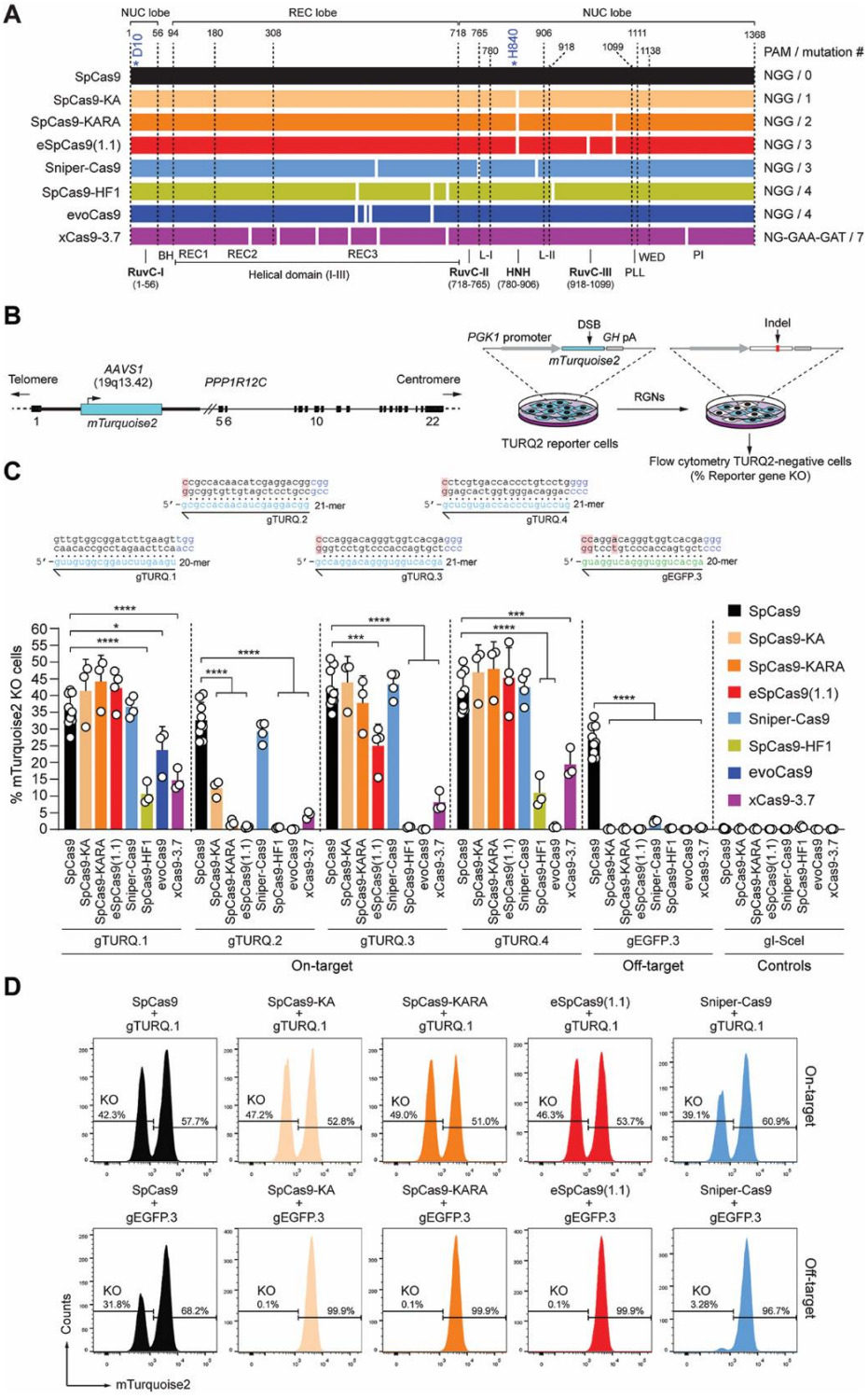


Figure 1. Comparing the activity and specificity of RGNs based on SpCas9 or SpCas9 variants. (**A**) Schematics of nucleases derived from the *S. Pyogenes* type II CRISPR system. Protein domains and mutations (white bars) are indicated. HNH, histidine-asparagine-histidine nuclease domain; RuvC, RNase H-like fold nuclease domain formed by tripartite assembly of RuvC-I, -II and -III. The HNH and RuvC domains in the nuclease lobe digest the target and non-target DNA strands, respectively. L-I and L-II, linker region I and II, respectively. Numerals correspond to the amino acid positions delimiting the various protein domains and motifs. BH, Arginine-rich bridge helix that connects the NUC and REC lobes; CTD, C-terminal domain in which the PAM-interacting motif (PI) is lodged; NUC and REC, nuclease and recognition lobes, respectively; PLL, phosphate lock loop. Asterisks mark residues D10 and H840 crucial for RuvC and HNH catalytic activities, respectively. The diagram of the *S. aureus* Cas9 (SaCas9) nuclease ortholog is also shown. (**B**) Gene knockout assays. TURQ2 cells contain an *mTurquoise2* transgene at intron 1 of *PPP1R12C* (*AAVS1* locus). Small insertions and deletions (indels) resulting from the action of programmable nucleases and NHEJ pathways at *mTurquoise2* yield gene knockouts quantifiable by flow cytometry. (**C**) Determining RGN activities. TURQ2 cells were transfected with plasmids expressing the indicated RGN components. The gRNAs gTURQ.1 through gTURQ.4 have spacers fully complementary to *mTurquoise2* sequences (on-target); *EGFP*-specific gEGFP.3 has a spacer with mismatches to a *mTurquoise2* sequence (off-target). The non-targeting gRNA gI-SceI was used as a negative control. Non-hybridizing DNA-gRNA bases are highlighted in red. Gene knockout frequencies were determined at 10 days post-transfection through flow cytometry of mTurquoise2-negative cells. Data are presented as mean ± S.D. of at least three independent biological replicates. Significant differences between datasets were calculated with one-way ANOVA followed by Dunnett's test for multiple comparisons; *0.01 < P < 0.05; ***0.0001 < P < 0.001; ****P < 0.0001. (**D**) Examples of gene knockout datasets. Histograms corresponding to TURQ2 cell populations subjected to RGNs with spacers complementary and partially complementary to a target sequence (top and bottom panels, respectively).

Functional screens identify a versatile set of high-specificity nickases

After comparing SpCas9 nuclease performances, we generated isogenic constructs expressing the corresponding RuvC-disabled nicking forms; SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A}, Sniper-Cas9^{D10A}, SpCas9-HF1^{D10A}, evoCas9^{D10A} and xCas9-3.7^{D10A} (**Figure 2A**). These enzymes were subsequently screened in quantitative assays as dual nRGNs for establishing their gene knockout activities upon simultaneous SSB formation. These assays were initiated by exposing H27 cells to dual nRGNs harboring the conventional SpCas9^{D10A} protein or each of the nicking variants coupled to different gRNA pairs (**Figure 2B**). The frequencies of gene knockouts resulting from the concerted action of nRGN pairs were measured through flow cytometry. Notably, these experiments showed that dual nRGNs containing SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A} or Sniper-Cas9^{D10A} can be as active as or more active than dual nRGNs built on the original SpCas9^{D10A} protein (**Figure 2B**). In contrast, dual nRGNs harboring SpCas9-HF1^{D10A}, evoCas9^{D10A} or xCas9-3.7^{D10A} were less active than their respective SpCas9^{D10A}-containing dual nRGN counterparts. Targeted deep sequencing analysis of 'footprints' induced by dual nRGNs containing the gRNA pair gEGFP.2/gEGFP.21 confirmed the flow cytometry data (**Figure 2B**) on their differential DNA cleavage activities (**Figure 2C** and **Supplementary Figure S2**). In most instances, this analysis further uncovered a clear preponderance of deletions over insertions and substitutions with a skewing of the deletions centred around the gEGFP.2 target site (**Figure 2C** and **Supplementary Figure S2**) which, of the two gRNAs, is the most effective when coupled to Cas9 nucleases (**Supplementary Figures S1** and **S3**). Interestingly, sequence profiling of the most frequent 'footprints' revealed a paucity of insertions in cells treated with dual nRGNs harbouring members of the nickase variant sub-set formed by the single, double and triple mutants SpCas9-KA, SpCas9-KARA and eSpCas9(1.1), respectively (**Figure 2C** and **Supplementary Figure S2B**). This data suggests that the choice of nickase variant impacts the complexity of dual nRGN-induced target DNA changes.

The best-performing dual nRGNs, i.e., those with SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A} or Sniper-Cas9^{D10A}, were less active when placed in a so-called PAM-in arrangement (**Figure 2B**). This data is in agreement with previous experiments using conventional dual nRGNs in which among PAM-out and PAM-in arrangements, the former normally yields higher DNA cleaving activities (39). Interestingly, not only for the original SpCas9^{D10A} nickase but also for each of the four best-performing SpCas9^{D10A} variants, the highest absolute frequencies of gene knockouts were detected in cultures exposed to the gRNA pair in which one of the members had a non-canonical 21-mer spacer (i.e. gEGFP.21) (**Figure 2B**). This result is especially notable for dual nRGNs containing eSpCas9(1.1)^{D10A} in that its parental eSpCas9(1.1) nuclease was poorly active when provided with gEGFP.21 but highly active when coupled to gEGFP.2 (**Supplementary Figures S1** and **S3**). This data suggests that in the context of dual nRGNs a highly active complex can rescue or compensate for a poorly active neighbouring complex. In particular, it is possible that non-canonical 21-mer spacers mostly affect the RuvC domain of eSpCas9(1.1) which is functionally absent in dual nRGNs with eSpCas9(1.1)^{D10A}. Finally, with the exception of xCas9-3.7 and xCas9-3.7^{D10A}, western blot analysis revealed similar amounts of cleaving and nicking SpCas9 enzymes and dual nRGNs in transfected cells (**Supplementary Figure**

S4). Importantly, dose-response experiments showed that gene knockout activities of RGNs and dual nRGNs containing xCas9-3.7 and xCas9-3.7^{D10A}, respectively, were not affected or scarcely affected by increasing the amounts of these proteins (**Supplementary Figure S5**).

Next, we sought to study the relationship between the activities and specificities of individual nRGNs endowed either with either SpCas9^{D10A} or each of the SpCas9^{D10A} variants. To detect targeted SSBs catalyzed by individual nRGNs, we established an assay based on delivering two types of SSB-inducing complexes into reporter cells. The first is a test *S. pyogenes* nRGN whose activity and specificity one wishes to determine; the second is a fixed S. aureus nRGN whose role is that of inducing a SSB off-set to that made by the test nRGN. Hence, this Cas9 orthogonal readout system permits sensitive and accurate measurements of nicking activities via recapitulating the modus operandi of dual nRGNs (**Figure 3A**, left panel). Crucially, by providing SpCas9^{D10A} variants with gRNAs presenting an array of mismatches to reporter sequences (**Figure 3A**, central panel), this readout system equally permits precisely assessing nRGN specificities which, as per definition, should inversely correlate with off-target nRGN activities (**Figure 3A**, right panel).

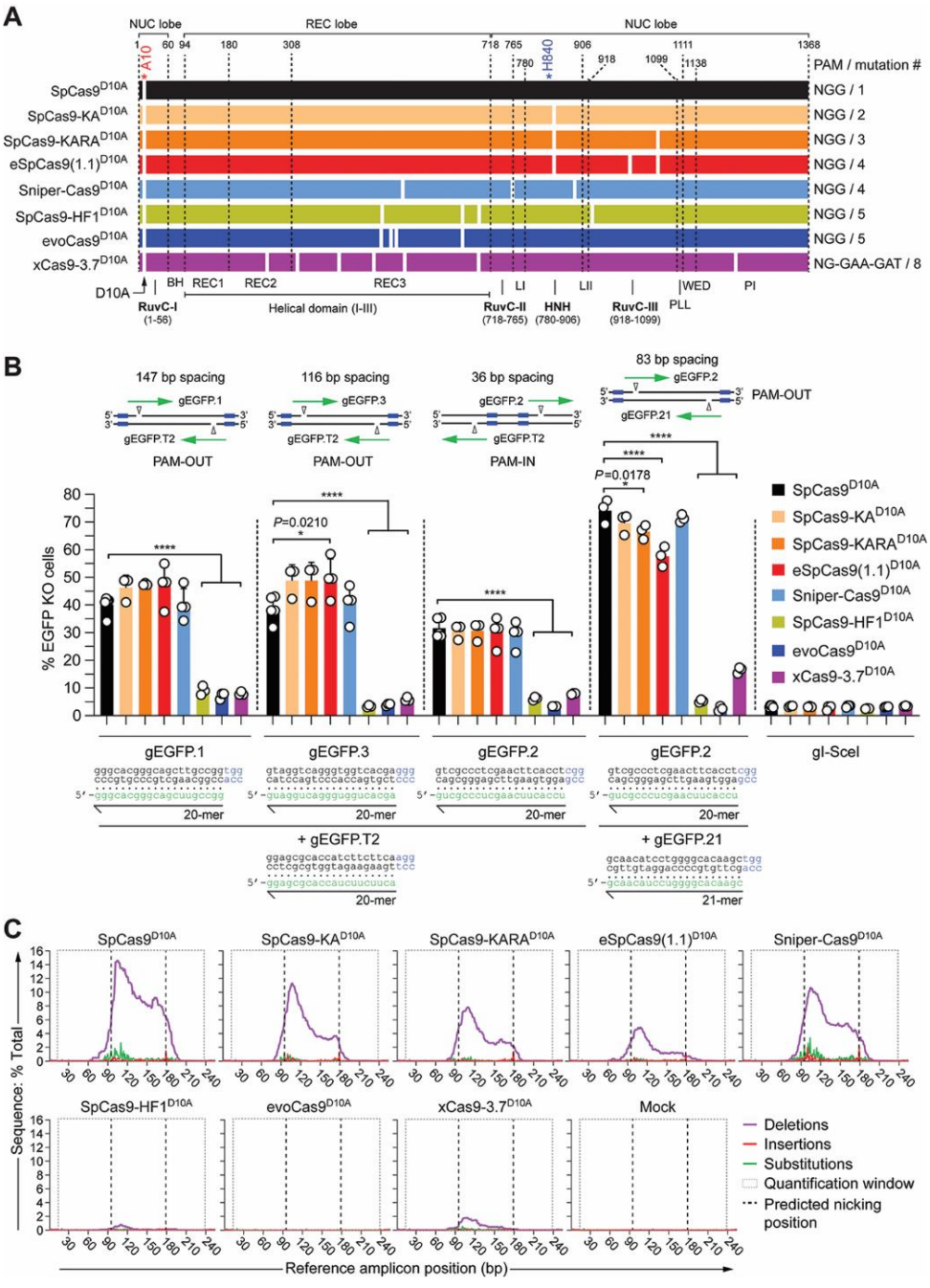


Figure 2. Comparing the activity of dual nRGNs based on SpCas9^{D10A} or SpCas9^{D10A} variants. (**A**) Schematics of original SpCas9^{D10A} and SpCas9^{D10A} variants generated for this study. Domains and mutations (white bars) in the nickases derived from the *S. pyogenes* type II CRISPR system are indicated. All nickases were obtained by introducing the RuvC-disabling D10A mutation into the nucleases depicted in **Figure 1A**. (**B**) Determining dual nRGN activities by gene knockout assays. EGFP-expressing H27 cells were transfected with constructs encoding the indicated dual nRGNs. Blue boxes, green arrows and open arrowheads in the insets indicate PAMs, gRNA spacers and nicking positions, respectively. Dual nRGNs with PAM-out and PAM-in arrangements were assessed. The non-targeting gRNA gI-SceI was used as a negative control. Gene knockout frequencies were determined by flow cytometry of EGFP-negative cells at 10 days post-transfection. Data are shown as mean ± S.D. of at least three independent biological replicates. Significance amongst the indicated datasets was calculated with one-way ANOVA followed by Dunnett's test for multiple comparisons; *0.01 < *P* < 0.05; ****$P < 0.0001$. (**C**) Characterization of dual nRGN "footprints by amplicon deep sequencing. H27 cells were exposed to dual nRGNs consisting of the indicated nickases loaded with gEGFP.2 and gEGFP.21. The types and frequencies of gene modifications detected at 48 hours post-transfection within the *EGFP* target sequence are plotted.

Previous experiments have indicated that RGN tolerance to DNA-gRNA mismatches roughly increases with the distance of these mismatches to the PAM (1,9). In keeping with these data, the 10–12 nts most proximal to the PAM have been proposed to constitute a 'seed region' in which DNA-gRNA mismatches are particularly detrimental for RGN activity (1,9). Hence, to increase the stringency of the nickase specificity screens in TURQ2 cells and maximize detecting differences in on-to-off target ratios (specificity indexes), we used a panel of gRNAs whose single, double and triple mismatches to reporter sequences were all located outside this 'seed region' (gOT-1 through gOT-10) (**Figure 3A**, central panel, **Supplementary Figure S6**). Furthermore, we chose to build the panel of mismatching gRNAs on basis of gEGFP.2 as its spacer is fully complementary to a *mTurquoise2* target site and led to comparably robust gene knockout frequencies irrespective of the SpCas9 nuclease used (**Supplementary Figure S1**). The *mTurquoise2*-specific S. pyogenes gEGFP.2 and its target site-mismatched derivatives were combined with a fixed fully-matching *S. aureus* gRNA (Sa-gRNA-G).

Consistent with the previous experiments using *S. pyogenes* gRNA pairs (**Figure 2**), gene knockout levels attained with gEGFP.2 and Sa-gRNA-G revealed that SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} constitute robust SSB-inducing enzymes (**Figure 3B**, compare respective first bars). Equally in line with the previous data (**Figure 2**), SpCas9-HF1^{D10A} and evoCas9^{D10A} were the least performing nickases whilst, in this case, xCas9-3.7^{D10A} presented an intermediate nicking activity (**Figure 3B**, compare respective first bars). Together, these data demonstrate a striking difference in the tolerability of high-specificity SpCas9 nucleases to the D10A mutation and, hence, to their conversion into operative nickases.

The specificity assays involving loading the different SpCas9^{D10A} nickases with gRNAs partially complementary to the gEGFP.2 target DNA, generically showed a mismatch number-dependent decrease in gene knockout frequencies (**Figure 3B** and **Supplementary Figure S7**). Among the high-activity nickases, i.e. SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, Sniper-Cas9^{D10A} and eSpCas9(1.1)^{D10A}, the latter was the most consistent in discriminating 1-nt, 2-nt and 3-nt gRNA–DNA mismatches, as indicated by the respective specificity indexes (**Figure 3C**). The high specificity of eSpCas9(1.1)^{D10A} was confirmed through gene knockout experiments using dual nRGNs exclusively with *S. pyogenes* gRNAs (**Supplementary Figure S8**). Amongst the low-activity nickases, i.e. SpCas9-HF1^{D10A} and evoCas9^{D10A}, the former outperformed the latter in that, besides presenting higher on-target activity (**Figure 3B**), it was generally better at discriminating 1-nt, 2-nt and 3-nt mismatches (**Figure 3C**). Finally, the intermediate-activity xCas9-3.7^{D10A} nickase had its highest discriminating power at gRNA–DNA sequences with 2-nt and 3-nt mismatches (**Figure 3C**). Despite their low activities, SpCas9-HF1^{D10A}, evoCas9^{D10A} and xCas9-3.7^{D10A} offer higher specificities than SpCas9^{D10A}. In fact, for gRNA–DNA heteroduplexes with 3-nt mismatches, xCas9-3.7^{D10A} presented specificity indexes superior to those of Sniper-Cas9^{D10A}, SpCas9-HF1^{D10A} and evoCas9^{D10A} (**Figure 3C**). Importantly, notwithstanding their varying on-target cleaving proficiencies, all engineered SpCas9^{D10A} variants were shown to be more specific than their parental SpCas9^{D10A} counterpart (**Figure 3B** and **C**).

We conclude that these reagents form a broad and versatile set of RNA-programmable nicking enzymes whose activities and/or specificities are superior to those of the commonly used SpCas9^{D10A} protein.

Three-tier precision gene editing based on integrating high-specificity dual nicking RGN and truncated gRNA principles

Depending on their particular sequence, gRNAs with <20-mer spacers can significantly decrease SpCas9 off-target activities (40). It was postulated that amongst RGNs with 5'-truncated and full-length gRNAs, mismatches mostly destabilize the former leading to higher specificities (40). Hence, coupling high-specificity SpCas9 nucleases to validated truncated gRNAs is an appealing two-tier strategy to further reduce RGN off-target activities. Yet, similarly to 5' non-hybridizing and extended gRNAs (36-38), truncated gRNAs can significantly hamper the on-target activities of high-specificity SpCas9 nucleases (23–25,31). To investigate a multi-tier approach for maximizing gene-editing tool precision based on integrating high-specificity dual nRGN and truncated gRNA principles, we tested the effect of truncated gRNAs on the activities of RGNs and dual nRGNs with high-specificity cleaving and nicking SpCas9 enzymes, respectively. To this end, H27 cells were subjected to dual nRGNs formed by gRNA pairs in which both members were full-length (i.e. gEGFP7/gEGFP6.FL20) (**Figure 4A**, open bars in top graphs) or one member was full-length and the other was truncated (i.e. gEGFP7/gEGFP6.tru19 and gEGFP7/gEGFP6.tru17) (**Figure 4A**, open bars in bottom graphs). As references, H27 reporter cells were exposed to RGNs with full-length gRNAs (i.e. gEGFP7 and gEGFP6.FL20) (**Figure 4A**, solid bars in top graphs) or truncated gRNAs (i.e. gEGFP6.tru19 and gEGFP6.tru17) (**Figure 4A**, solid bars in bottom graphs).

The cumulative gene knockout experiments revealed that the Sniper-Cas9 nuclease was the variant most compatible with truncated gRNAs with the 17-mer gRNA in particular only yielding gene knockouts once associated with this high-specificity nuclease (**Figure 4A**, solid cyan bar in bottom right-hand graph). These results are generically consistent with those of another study indicating that when compared to eSpCas9(1.1), SpCas9-HF1 and evoCas9, Sniper-Cas9 was least affected by 5'-end gRNA truncation (24). Crucially, nickases SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A}, once combined with gRNA pair gEGFP7/gEGFP6.tru17, invariably performed better than their respective high-specificity nucleases provided with gEGFP6.tru17 (**Figure 4A**, bottom right-hand graph). In fact, although the nucleases SpCas9-KA, SpCas9-KARA and eSpCas9(1.1) presented robust activities with gEGFP6.tru19, their activities were reduced to background levels once coupled to gEGFP6.tru17 (**Figure 4A**, compare respective solid bars in bottom graphs). Moreover, amongst the high-specificity dual nRGNs, those harboring Sniper-Cas9^{D10A} achieved the highest absolute levels of target gene knockout (**Figure 4A**, open bars in bottom right-hand graph). This conclusion was further supported through complementary experiments in which gene knockout levels induced by dual nRGNs with truncated gRNAs were measured against those triggered by dual nRGNs containing full-length gRNA pairs (**Figure 4B**). Additional experiments involving a Cas9 orthogonal readout system and gRNAs with 17-, 18- and 19-mer spacers confirmed that dual nRGNs based on Sniper-Cas9^{D10A} are compatible with truncated gRNAs (**Figure 4C**). Follow-up experiments using the same Cas9 orthogonal assay, established that Sniper-Cas9D10A endowed with truncated gRNAs can discriminate gRNA–DNA mismatches significantly better than SpCas9^{D10A} (**Figure 4D**). In fact, single base-pair mismatches located at PAM distal positions in 18-mer spacers sufficed to bring Sniper-Cas9^{D10A} nicking activities at near background levels (**Figure 4D**). Taken together, these data validate a three-tier precision gene editing strategy based on integrating into the dual nRGN concept, the high-specificity nickase and truncated gRNA principles.

Standard and high-specificity dual nRGN activities are comparable at heterochromatic target sites

The previous functional screens of standard and high-specificity nucleolytic enzymes, demonstrated that eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} offer a favourable and complementary set of attributes, as judged by their efficiency, specificity and versatility. In particular, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} display enhanced specificity and mostly retain the activity of SpCas9^{D10A}. The specificity of eSpCas9(1.1)^{D10A} is superior to that of Sniper-Cas9^{D10A}, yet Sniper-Cas9^{D10A} is more compatible with non-canonical gRNAs, including truncated gRNAs, than eSpCas9(1.1)^{D10A}.

We thus progressed by investigating these nickases further, starting with their performance at alternate higher-order chromatin conformations. It is known that compact heterochromatic states can hinder gene-editing tool activities, including those of transcription activator-like effector nucleases, RGNs and standard dual nRGNs (30,31). To compare standard and high-specificity dual nRGNs at isogenic target sites packed in loose euchromatin versus compact heterochromatin, we employed HEK.EGFP$^{TetO.KRAB}$ reporter cells (30). These cells allow for doxycycline-dependent control over Krüppel-associated box

(KRAB)-mediated recruitment of endogenous epigenetic remodelling complexes to programmable nuclease target sites (**Figure 5A** and **Supplementary Figure S9A**). These complexes consist of, among other factors, KRAB-Associated Protein 1 (KAP1) and heterochromatin protein 1 (HP1) (**Figure 5A**). As expected, dual nRGNs based on SpCas9^{D10A}, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} were all significantly more active at euchromatic sequences in doxycycline-treated HEK.EGFP$^{TetO.KRAB}$ cells than at the same heterochromatic sequences in untreated HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 5B**, **C** and **D**, respectively). Importantly, at KRAB-impinged heterochromatin, high-specificity dual nRGNs containing Sniper-Cas9^{D10A} or eSpCas9)1.1)^{D10A} performed similarly to standard dual nRGNs (**Figure 5E** and Supplementary **Figure S9B**).

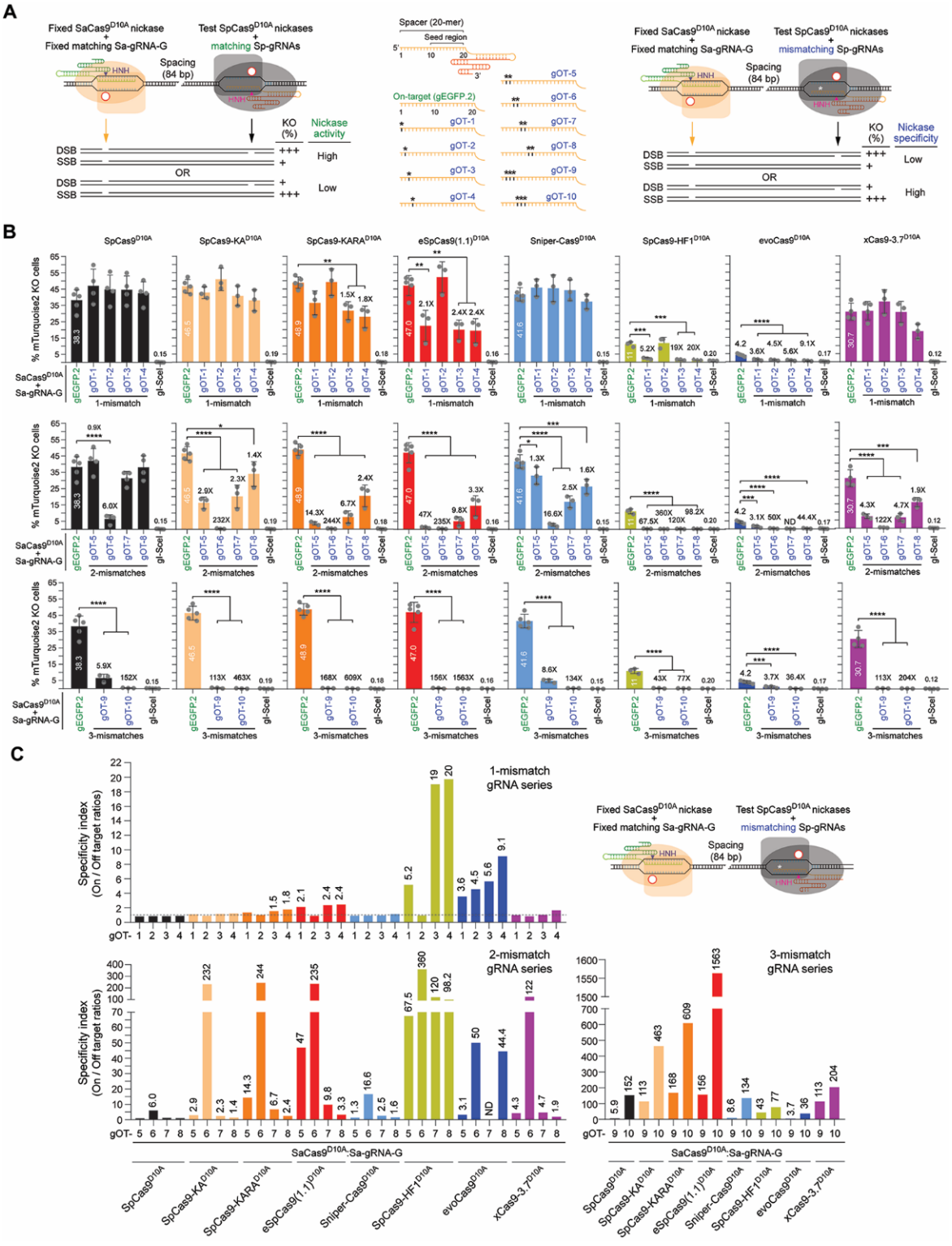


Figure 3. Comparing the performance of nRGNs based on SpCas9^{D10A} or SpCas9^{D10A} variants. (**A**) Cas9 orthogonal assay for determining the activity and specificity of nRGNs. A fixed *S. aureus* nRGN (orange) is introduced together with a test *S. pyogenes* nRGN (black) into reporter cells. Coordinated formation of SSBs at opposite strands of a bipartite reporter-encoding sequence by each nicking complex results in DSB-induced gene knockouts. Comparing the activities and specificities of different nickases can be assessed by loading *S. pyogenes* gRNAs with fully or partially hybridizing spacers (left and central panel, respectively). Test nRGN activities and specificities are directly and inversely proportional, respectively, to gene knockout frequencies (right panel). The fully matching spacer of *S. pyogenes* gEGFP.2 is drawn in relation to *S. pyogenes* gRNA spacers with 1-nt, 2-nt or 3-nt mismatches (asterisks) outside the seed region (central panel). (**B**) Comparing the specificity profiles of nRGNs with different nickases. Reporter cells were transfected with plasmids encoding the denoted nRGNs. The spacers of the three sets of off-target (OT) gRNAs, i.e., gOT-1 through gOT-4, gOT-5 through gOT-8 and gOT-9 plus gOT-10 have 1-nt, 2-nt and 3-nt mismatches, respectively, to the target sequence of gEGFP.2. Gene knockout levels were determined at 10 days post-transfection through flow cytometry of mTurquoise2-negative cells. Datasets correspond to mean ± S.D. of a minimum of three independent biological replicates. Significance between the indicated datasets was calculated with one-way ANOVA followed by Tukey's test for multiple comparisons; *0.01< *P* <0.05; **0.001< *P* <0.01; ***0.0001< *P* <0.001; ****P<0.0001. (**C**) The specificity indexes corresponding to DNA cleavage frequencies induced by nRGNs with *mTurquoise2*-matched gEGFP.2 divided by those triggered with *mTurquoise2*-mismatched gRNAs gOT-1 through gOT-10, are plotted. The statistically significant nRGN specificity indexes are presented above the respective bars.

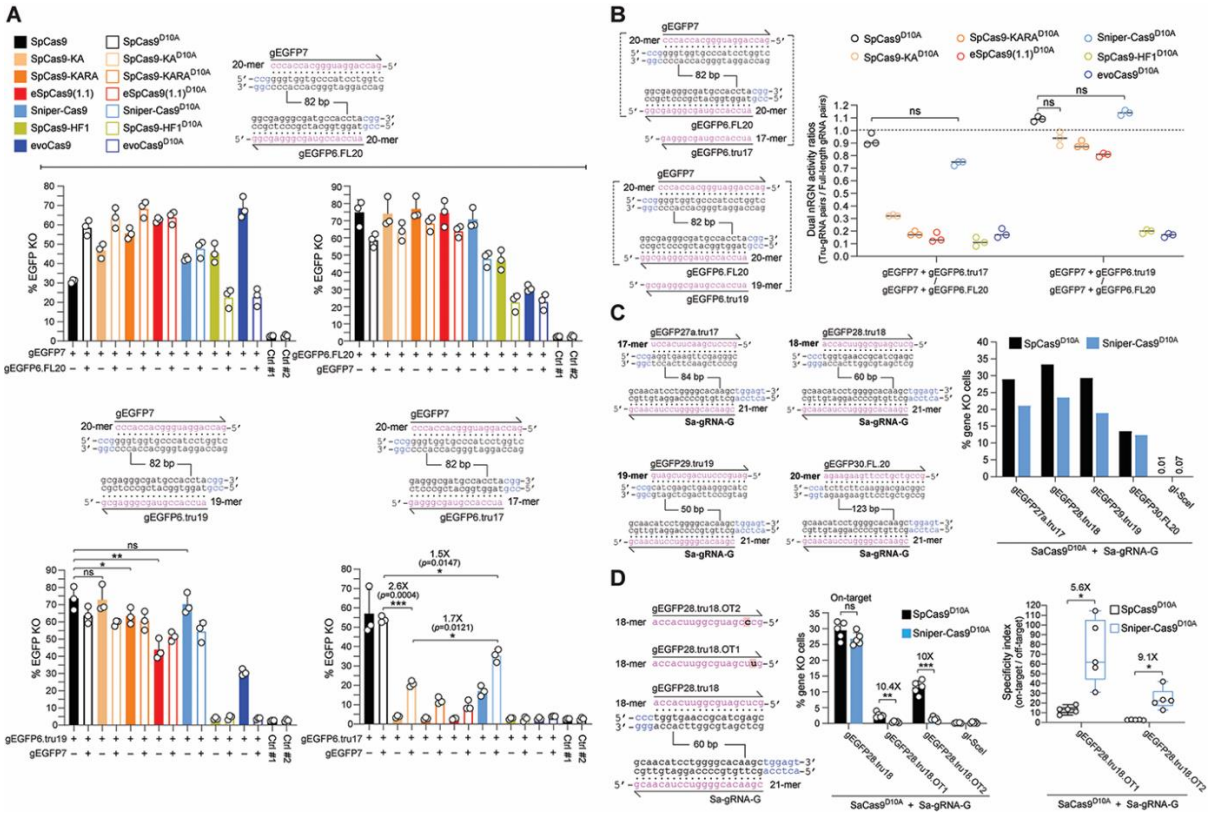


Figure 4. Investigating the integration of high-specificity dual nRGN and truncated gRNA principles. (**A**) Functional screening of high-specificity dual nRGNs with full-length and truncated gRNAs. EGFP-expressing H27 cells were exposed to dual nRGNs (open bars) containing a full-length gRNA pair (top panel) or expressing dual nRGNs harboring gRNA pairs with a truncated member (bottom panel). As references, H27 cells were exposed to RGNs (solid bars) with the same full-length gRNAs or truncated gRNAs. Results are presented as mean ± S.D. of independent biological replicates (n=3). Significance between the indicated datasets was calculated using two-tailed Student's t tests. *0.01 < P < 0.05; **0.001< P < 0.01; p≥0.05 was considered non-significant (ns). (**B**) Testing the effect of full-length versus truncated gRNAs on dual nRGN activities. Dual RGN activity ratios obtained by dividing DNA cleavage frequencies induced with gRNA pairs containing a truncated member by those triggered with gRNA pairs with full-length gRNAs (panel **A**). Data are shown as mean ± S.D. of independent biological replicates (n=3). Significance between the indicated datasets was calculated by one-way ANOVA followed by Dunnett's test for multiple comparisons; p≥0.05 was considered non-significant (ns). (**C**) Assessing the activities of nRGNs with truncated gRNAs. The *S. aureus* SaCas9:Sa-gRNA-G complex was introduced into TURQ2 cells together with *S. pyogenes* complexes formed by SpCas9^{D10A} or Sniper-Cas9^{D10A} loaded with 17-, 18-, 19- or 20-mer gRNAs specific for *EGFP* and *mTurquoise2* sequences. The frequencies of SSBs induced by each of the *S. pyogenes* nRGNs were established by flow cytometry of mTurquoise2-negative cells. (**D**) Testing the specificities of nRGNs with truncated gRNAs. The *S. aureus* SaCas9:Sa-gRNA-G complex was delivered into TURQ2 cells together with *S. pyogenes* complexes formed by SpCas9^{D10A} or Sniper-Cas9^{D10A} coupled to 18-mer spacer gRNAs specific for *EGFP* and *mTurquoise2* sequences with no mismatches or with a single mismatch (red boxes) to a transgene sequence. PAMs for *S. pyogenes* and *S. aureus* Cas9 proteins are highlighted in blue (left panel). *S. pyogenes* nRGN activities were determined by mTurquoise2-negative cell quantification, with SpCas9^{D10A} showing significantly more tolerance to gRNA-DNA mismatches than Sniper-Cas9^{D10A} as presented in absolute and relative terms (graphs in middle and right panels, respectively). In the middle panel, the data are presented as mean ± S.D. of independent biological replicates (n=5). Significance between the indicated datasets was calculated with two-tailed Student's t tests. **0.001< P < 0.01; ***0.0001 < P < 0.001; P ≥

0.05 was considered non-significant (ns). In the right panel, Box plot of independent biological replicates (n=5), with significances calculated through two-tailed Student's t tests; *0.01 < P < 0.05. In all experimental settings, gene knockout levels, were determined by flow cytometry of mTurquoise2-negative cells at 10 days post-transfection.

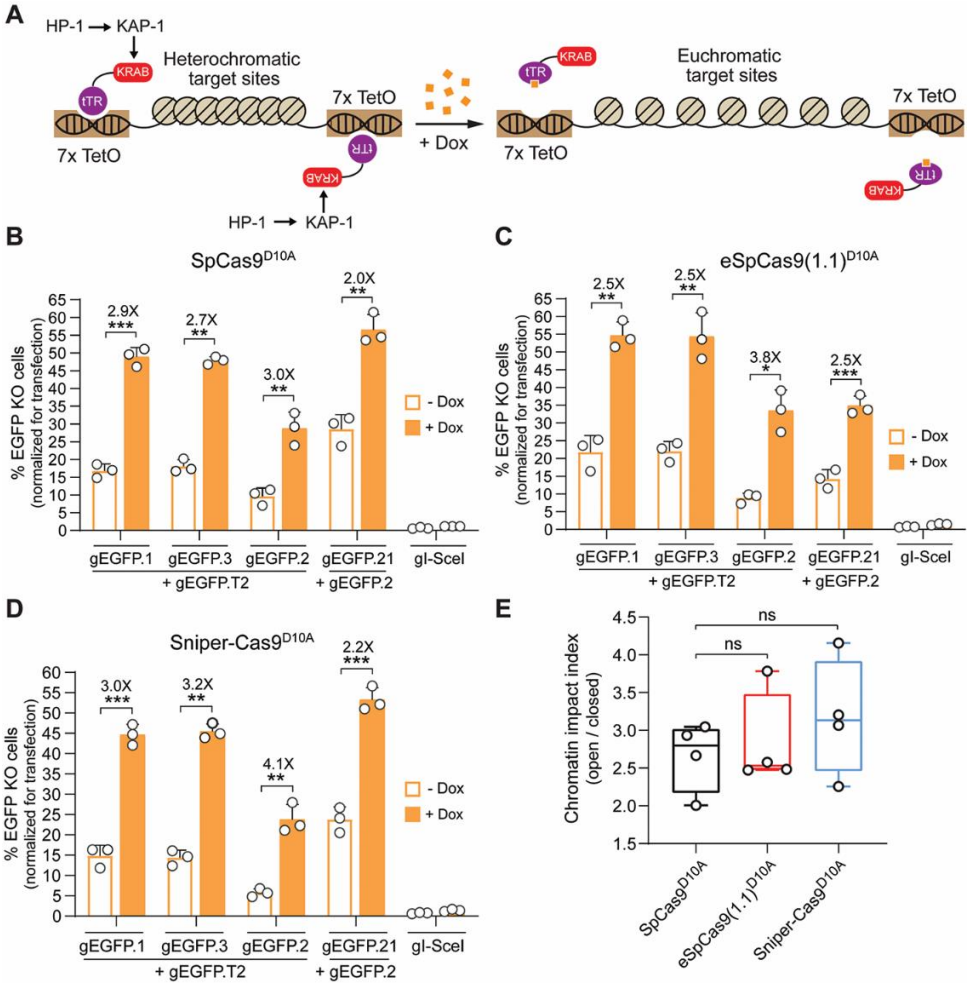


Figure 5. Comparing standard versus high-specificity dual nRGNs at alternate chromatin states. (**A**) Diagram of the experimental system. Doxycycline (Dox) availability regulates higher-order chromatin conformations that are controlled by KRAB-mediated recruitment of cellular silencing complexes to target sequences. In the absence of Dox, the tTR-KRAB fusion protein binds to *TetO* elements leading to the nucleation of cellular epigenetic modulators (e.g. KAP1 and HP1) and ensuing formation of compact heterochromatin at *EGFP* target sequences. In the presence of Dox, tTR-KRAB cannot bind to DNA, resulting in the maintenance of a relaxed euchromatin conformation at the same sequences. HEK.EGFP$^{TetO.KRAB}$ cells treated or not treated with Dox were subjected to the indicated sets of gene-editing reagents that differed through their inclusion of either SpCas9^{D10A} (**B**), eSpCas9(1.1)^{D10A} (**C**) or Sniper-Cas9^{D10A} (**D**). After eliminating gene-editing reagents by sub-culturing and exposing both culture types to Dox, to assure transgene expression, *EGFP* knockout frequencies were determined by flow cytometry (**Supplementary Figure S9A**). Data are presented as mean ± S.D. of independent biological replicates (n=3). Significance between datasets was calculated by two-tailed Student's t tests; *0.01 < P < 0.05; **0.001< P < 0.01; ***0.0001 < P < 0.001. (**E**) Cumulative chromatin impact indexes. Box plot presenting the chromatin impact indexes obtained by dividing gene knockout mean frequencies determined in the presence and absence of Dox (solid and open bars, respectively) (Figure S9b). Significance between the data points was calculated by one-way ANOVA followed by Dunnett's test for multiple comparisons; $P \geq 0.05$ was considered non-significant (ns).

The previous functional screens of standard and high-specificity nucleolytic enzymes, demonstrated that eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} offer a favourable and complementary set of attributes, as judged by their efficiency, specificity and versatility. In particular, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} mostly retain the activity of SpCas9^{D10A} while displaying enhanced specificity. While the specificity of eSpCas9(1.1)^{D10A} is superior to that of Sniper-Cas9^{D10A}, Sniper-Cas9^{D10A} is more compatible with non-canonical gRNAs, including truncated gRNAs, than eSpCas9(1.1)^{D10A}.

We thus progressed by investigating these nickases further, starting with their performance at alternate higher-order chromatin conformations. It is known that compact heterochromatic states can hinder gene-editing tool activities, including those of transcription activator-like effector nucleases, RGNs and

standard dual nRGNs (30,31). To compare standard and high-specificity dual nRGNs at isogenic target sites packed in loose euchromatin versus compact heterochromatin, we employed HEK.EGFP$^{TetO.KRAB}$ reporter cells (30). These cells allow for doxycycline-dependent control over Krüppel-associated box (KRAB)-mediated recruitment of endogenous epigenetic remodelling complexes to programmable nuclease target sites (**Figure 5A** and **Supplementary Figure S9A**). These complexes consist of, among other factors, KRAB-Associated Protein 1 (KAP1) and heterochromatin protein 1 (HP1) (**Figure 5A**). As expected, dual nRGNs based on SpCas9^{D10A}, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A} were all significantly more active at euchromatic sequences in doxycycline-treated HEK.EGFP$^{TetO.KRAB}$ cells than at the same heterochromatic sequences in untreated HEK.EGFP$^{TetO.KRAB}$ cells (**Figures 5B**, **5C** and **5D**, respectively). Importantly, at KRAB-impinged heterochromatin, high-specificity dual nRGNs containing Sniper-Cas9^{D10A} or eSpCas9)1.1)^{D10A} performed similarly to standard dual nRGNs (**Figure 5E, Supplementary Figure S9B**).

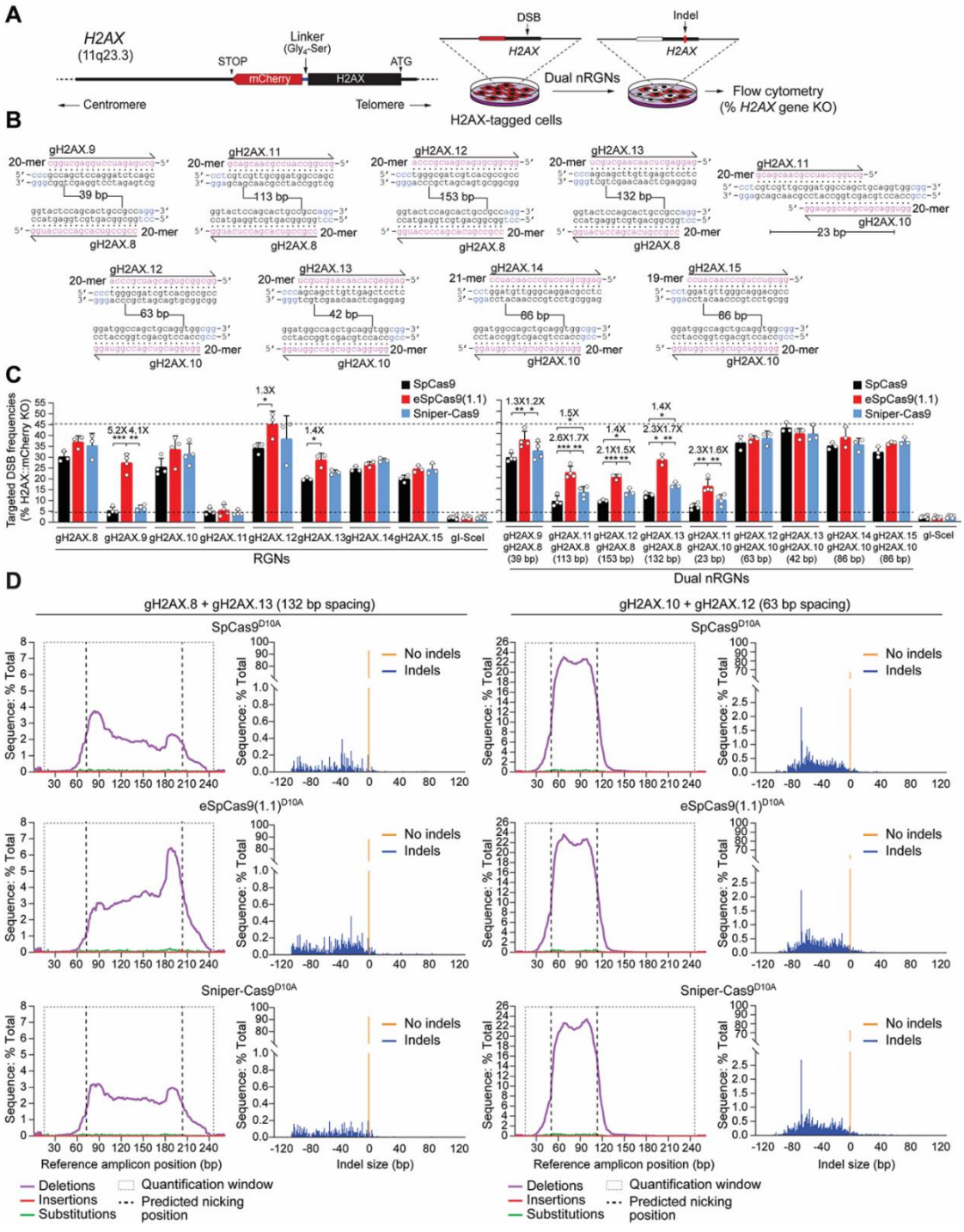


Figure 6. Testing the activity of RGNs and dual nRGNs at human genomic DNA. (**A**) Schematics of readout system. HeLa cells containing the *H2AX* gene in-frame with a *mCherry* reporter are exposed to dual nRGN components. Target DNA cleavage is assessed through flow cytometric quantification of mCherry-negative cells resulting from DSB-induced indels at *H2AX* sequences. (**B**) *H2AX*-targeting gRNAs. The gRNA spacer nucleotides are drawn annealing to the respective target DNA strands. PAM nucleotides are highlighted in blue. Numbers within broken line correspond to the spacing between gRNA pair members using as reference the base pair positions at which nicking occurs. (**C**) Functional screening of RGNs and dual nRGNs with standard or variant SpCas9 proteins at *H2AX*. H2AX::mCherry^{+} HeLa cells were transfected with plasmids expressing the indicated combinations of RGN and dual nRGN elements. DNA cleaving activities were assessed through flow cytometry of mCherry-negative cells at 10 days post-transfection. Dashed lines, corresponding to the lowest and highest DNA cleaving frequencies measured. Data are presented as mean ± S.D. of at least three independent biological replicates. Significance between the indicated datasets was calculated by two-tailed Student's *t* tests; *$0.01 < P < 0.05$; **$0.001< P < 0.01$; ***$0.0001 < P < 0.001$. (**D**) Characterization of dual nRGN "footprints" at *H2AX* alleles. The types and frequencies of gene modifications within the indicated dual nRGN target sequences were determined at 48 hours post-transfection by amplicon deep sequencing of HEK293T cells.

High-specificity dual nRGNs outperform standard dual nRGNs at genomic sequences

To compare the activities and specificities of dual nRGNs based on standard versus high-specificity nickases at endogenous genomic DNA, we targeted *H2AX* alleles in-frame with a mCherry reporter in HeLa cells. This set-up allows for sensitive flow cytometric quantification of DNA cleaving activities (**Figure 6A**). In initial experiments, SpCas9, eSpCas9(1.1), Sniper-Cas9, and their respective nicking derivatives, were used together with a panel of eighteen gRNAs (**Figure 6B**). In line with earlier results (**Figure 2**, **Supplementary Figures S1** and **S3**) (22), it was observed that low to intermediate RGN cleaving activities conferred by certain gRNAs can be bypassed via combining these gRNAs with a nickase and a second gRNA addressed to an off-set sequence; thus, effectively forming an operational dual nRGN complex (**Figure 6C**, compare left and right graphs). Most importantly, amidst the nine randomly selected PAM-out gRNA pairs covering a wide range of spacing lengths (**Figure 6B**), five yielded significantly higher *H2AX* knockout frequencies when combined with eSpCas9(1.1)^{D10A} instead of SpCas9^{D10A} (**Figure 6C**, right graph). Albeit to a lesser extent than eSpCas9(1.1)^{D10A}, three out of the nine gRNA pairs performed also better with Sniper-Cas9^{D10A} than with SpCas9^{D10A} (**Figure 6C**, right graph). Moreover, four gRNA pairs led to similar *H2AX* knockout frequencies, independently of the nickase to which they were joined (**Figure 6C**, right graph). These data indicate that dual nRGNs based on eSpCas9(1.1)^{D10A} can outperform SpCas9^{D10A}-containing dual nRGNs in inducing target DNA cleavage.

Targeted deep sequencing analysis of HEK293T cells exposed to dual nRGNs containing gRNA pairs gH2AX.8/gH2AX.13 and gH2AX.10/gH2AX.12, was consistent with the relative gene knockout levels measured by flow cytometry of HeLa reporter cells treated with the same gene-editing reagents (**Figure 6D** and **Supplementary Figure S10A**). This analysis further uncovered a vast representation of deletions over insertions and substitutions. In fact, sequence profiling revealed neither insertions nor substitutions amongst the ten most frequent 'footprints' (**Supplementary Figure S10B** and **S10C**). Interestingly, deletions triggered by dual nRGNs with the most spaced gRNAs (i.e. gH2AX.8/gH2AX.13) were often centred around either one of the target sites (**Supplementary Figure S10B**); whereas deletions induced by dual nRGNs with the least spaced gRNAs (i.e. gH2AX.10/gH2AX.12) mostly encompassed the intervening sequence (**Supplementary Figure S10C**). This data suggests that gRNA spacing impacts the complexity of dual nRGN-induced target DNA changes.

To strictly challenge the specificity of dual nRGNs based on SpCas9^{D10A}, eSpCas9(1.1.)^{D10A} and Sniper-Cas9^{D10A}, we next designed gRNAs bearing single nt mismatches to *H2AX* sequences mapping at PAM distal positions. HeLa cells expressing mCherry-tagged H2AX were exposed to dual nRGNs formed by gRNAs in which both or only one of their spacers contained 1-nt mismatches to *H2AX* sequences (**Figure 7**, top and bottom panels, respectively). In agreement with previous results (**Figure 3B** and **C**, **Supplementary Figure S8B** and **C**), these DNA cleaving specificity assays revealed that, amongst dual nRGNs based on SpCas9^{D10A}, Sniper-Cas9^{D10A} and eSpCas9(1.1.)^{D10A}, the latter are the most robust in discriminating subtle gRNA–DNA mismatches (**Figure 7**). This conclusion was strengthened through complementary experiments in which gene knockout levels triggered by dual nRGNs with DNA mismatching gRNAs were measured against those induced by dual nRGNs containing the respective, fully matching, gRNAs (**Figure 8**). We conclude that dual nRGNs based on eSpCas9(1.1)^{D10A} are valuable gene-editing tools in that they can outperform standard dual nRGNs at both the activity and specificity levels.

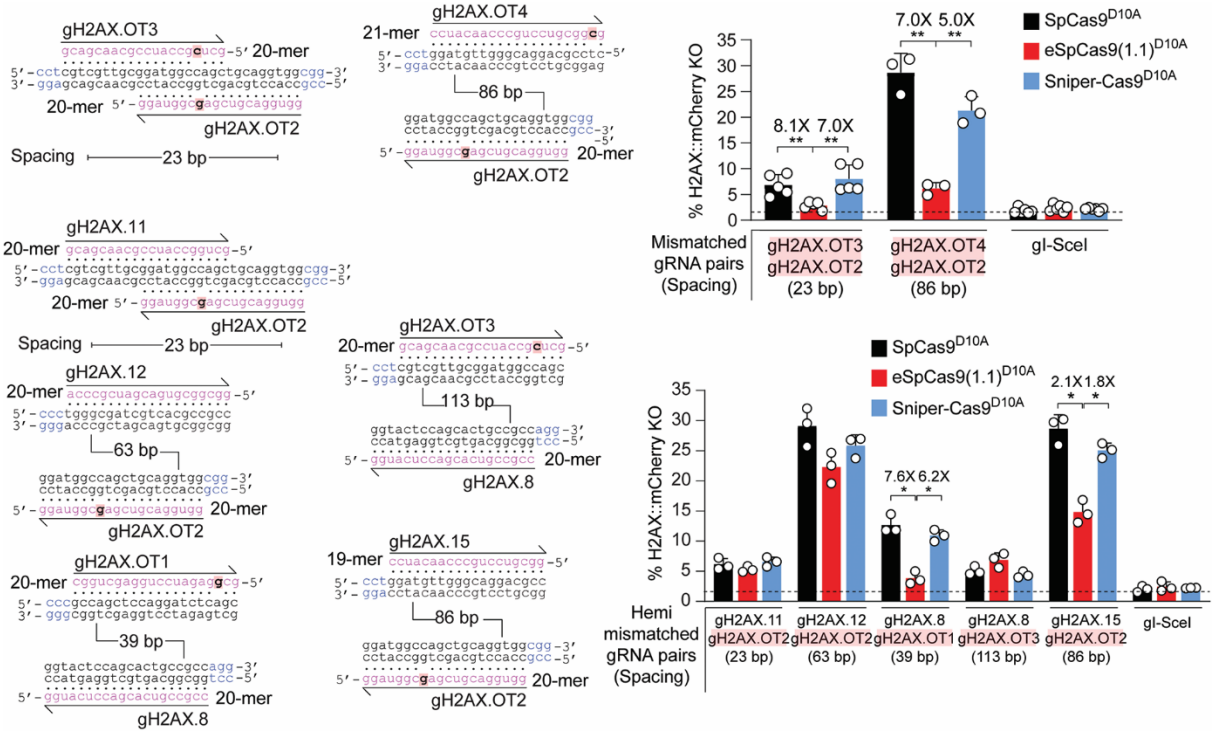


Figure 7. Testing the specificity of dual nRGNs at human genomic DNA. Specificity assay comparing standard and variant dual nRGNs containing gRNAs with mismatches to *H2AX* in both spacers (mismatched gRNA pairs) or only in one of the two spacers (hemi-mismatched gRNA pairs). *H2AX* and gRNA spacer sequences are drawn hybridizing to each other with mismatched and PAM nucleotides highlighted in red boxes and blue lettering, respectively. In these assays, the DNA mismatch discriminating power (specificity) of individual dual nRGNs inversely correlates with *H2AX* gene knockout frequencies. H2AX::mCherry^{+} HeLa cells were transfected with constructs expressing the denoted dual nRGNs. *H2AX* gene knockout frequencies were determined by flow cytometry of mCherry-negative cells at 10 days post-transfection. The results are expressed as mean ± S.D. of a minimum of three independent biological replicates. Significance between the indicated datasets was calculated by two-tailed Student's *t* tests; *0.01 < *P* < 0.05; **0.001< *P* < 0.01.

High-specificity dual nRGN "tiptoeing" achieves selective cleavage of genomic sites with high similarity to off-target sequences

OCT4 (a.k.a. *POU5F1*) is a coveted gene editing target owing to its essentiality for the maintenance of embryonic stem cells (ESCs) as well as for the maintenance and generation of iPSCs through cellular reprogramming (41,42). *OCT4* is equally crucial during early human embryogenesis (43). The selective modification of *OCT4* though programmable nucleases is, however, challenging due to the presence of *OCT4* pseudogenes in different chromosomes. Moreover, off-target sites located in *OCT4* pseudogenes combined with the particularly high sensitivity of pluripotent stem cells (PSCs) to few DSBs (44–46), renders the isolation of *OCT4*-edited PSCs highly inefficient (14,47,48). Indeed, OCT4 tagging experiments in PSCs involving recombination between target and pDonor^{OCT4} sequences (**Figure 9A**) triggered with TALENs (47) or RGNs (48) retrieved, respectively, no iPSC (n = 48) or only eight ESC (n = 288) clones that were correctly edited. Thus, to compare the capacity of standard and high-specificity dual nRGNs to distinguish target DNA from highly similar off-target genomic sequences, we performed HDR-mediated gene knock-in experiments at *OCT4* using pDonor^{OCT4} (**Figure 9A**). In particular, we asked whether the heightened single base-pair resolution of high-specificity dual nRGNs permits discriminating highly similar genomic sequences from each other by 'tiptoeing' over preexisting indels or single nucleotide polymorphisms (SNPs). To this end, HeLa cells were first transfected with pDonor^{OCT4} mixed with constructs encoding a panel of dual nRGNs based on SpCas9^{D10A} or eSpCas9(1.1)^{D10A} (**Figure 9A** and **B**). Colony-formation assays revealed that the number of cells acquiring puromycin resistance varied as a function of the nickase and gRNA pair used (**Figure 9B**). Most importantly, off-target analysis of genomic DNA from puromycin-resistance HeLa cell populations revealed that dual nRGNs with eSpCas9(1.1)^{D10A} were substantially more specific than their SpCas9^{D10A}-containing counterparts (**Figure 9C**). Indeed, six out of seven gRNA pairs readily led to DSB formation at *POU5F1P4* when coupled to SpCas9^{D10A}, whilst only two of these gRNA pairs induced DSBs at this locus once linked to eSpCas9(1.1)^{D10A} (**Figure 9C**, left panel). At *POU5F1P5*, out of eight gRNA pairs tested, two and one yielded off-target cleavage when coupled to SpCas9^{D10A} and eSpCas9(1.1)^{D10A},

respectively (**Figure 9C**, right panel). The fact that *POU5F1P4* and *POU5F1P5* overlap with coding genes (i.e. *ASH1L* and *HERC4*, respectively) further compounds the genotype of cells suffering off-target DSBs at these loci (**Supplementary Figure S11**). Moreover, clonal analysis assessing gene knock-ins at *OCT4* and pseudogene loci, established that the specificity of HDR-mediated gene editing was substantially higher (13-fold) when dual nRGNs were endowed with eSpCas9(1.1)^{D10A} instead of SpCas9^{D10A} (**Figure 9D** and **Supplementary Figure S12**). In particular, from 30 randomly selected HeLa cell clones derived from cultures exposed to pDonor^{OCT4} and SpCas9^{D10A}-based dual nRGNs, only 1 was properly edited, i.e., was targeted at *OCT4* (**Figure 9D**, top panels green arrow) and lacked mistargeted insertions at *OCT4* pseudogenes (**Supplementary Figure S12**). In contrast, 10 out of 23 clones isolated from cultures treated with pDonor^{OCT4} and eSpCas9(1.1)^{D10A}-based dual nRGNs, were properly edited (**Figure 9D**, bottom panels green arrows). Thus, although dual nRGNs are prevalently used for NHEJ-mediated gene knockouts, their capacity to induce HDR-mediated gene knock-ins broadens their applicability, especially if built on high-specificity nickases. Indeed, this data indicates that NHEJ- and HDR-based gene editing with dual nRGNs harboring eSpCas9(1.1)^{D10A} permits a more judicious access to specific genomic variants through 'tiptoeing' over short preexisting polymorphisms.

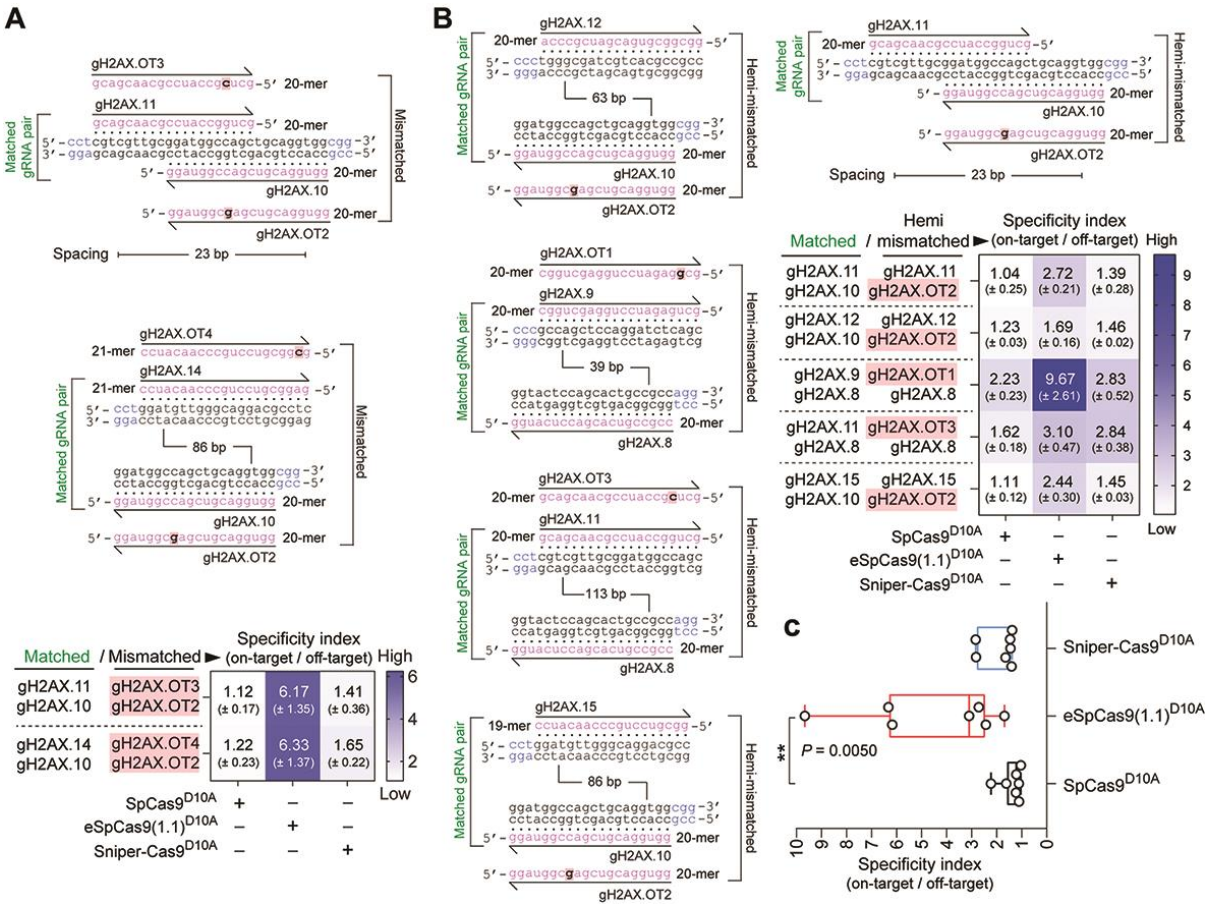


Figure 8. Testing the effect of sequence mismatches on standard and high-specificity dual nRGN activities. (**A**) Comparing standard versus variant dual nRGNs with DNA-mismatched gRNA pairs. Dual nRGNs based on SpCas9^{D10A}, eSpCas9(1.1)^{D10A} or Sniper-Cas9^{D10A}, coupled to *H2AX*-matched or mismatched gRNA pairs, were introduced into H2AX::mCherry^{+} HeLa cells. The heatmap presents dual nRGN specificity indexes (mean ± S.D.) resulting from dividing the gene knockout frequencies induced with *H2AX*-matched gRNA pairs by those attained with the respective mismatched gRNA pairs. (**B**) Comparing standard versus variant dual nRGNs with DNA hemi-mismatched gRNA pairs. Dual nRGNs based on SpCas9^{D10A}, eSpCas9(1.1)^{D10A} or Sniper-Cas9^{D10A}, linked to *H2AX*-matched or hemi-mismatched gRNA pairs, were delivered into H2AX::mCherry^{+} HeLa cells. The heatmap depicts dual nRGN specificity indexes (mean ± S.D.) derived from dividing the gene knockout frequencies achieved with *H2AX*-matched gRNA pairs by those attained with the respective hemi-mismatched gRNA pairs. (**C**) Cumulative specificity indexes. Box plot of the specificity indexes presented in the heatmaps of panels A and B. In all experimental settings, gene knockout levels, corresponding to at least three independent biological replicates, were determined by flow cytometry of EGFP-negative cells at 10 days post-transfection. Significance between datasets was calculated with one-way ANOVA followed by Dunnett's test for multiple comparisons; ** 0.001< P < 0.01.

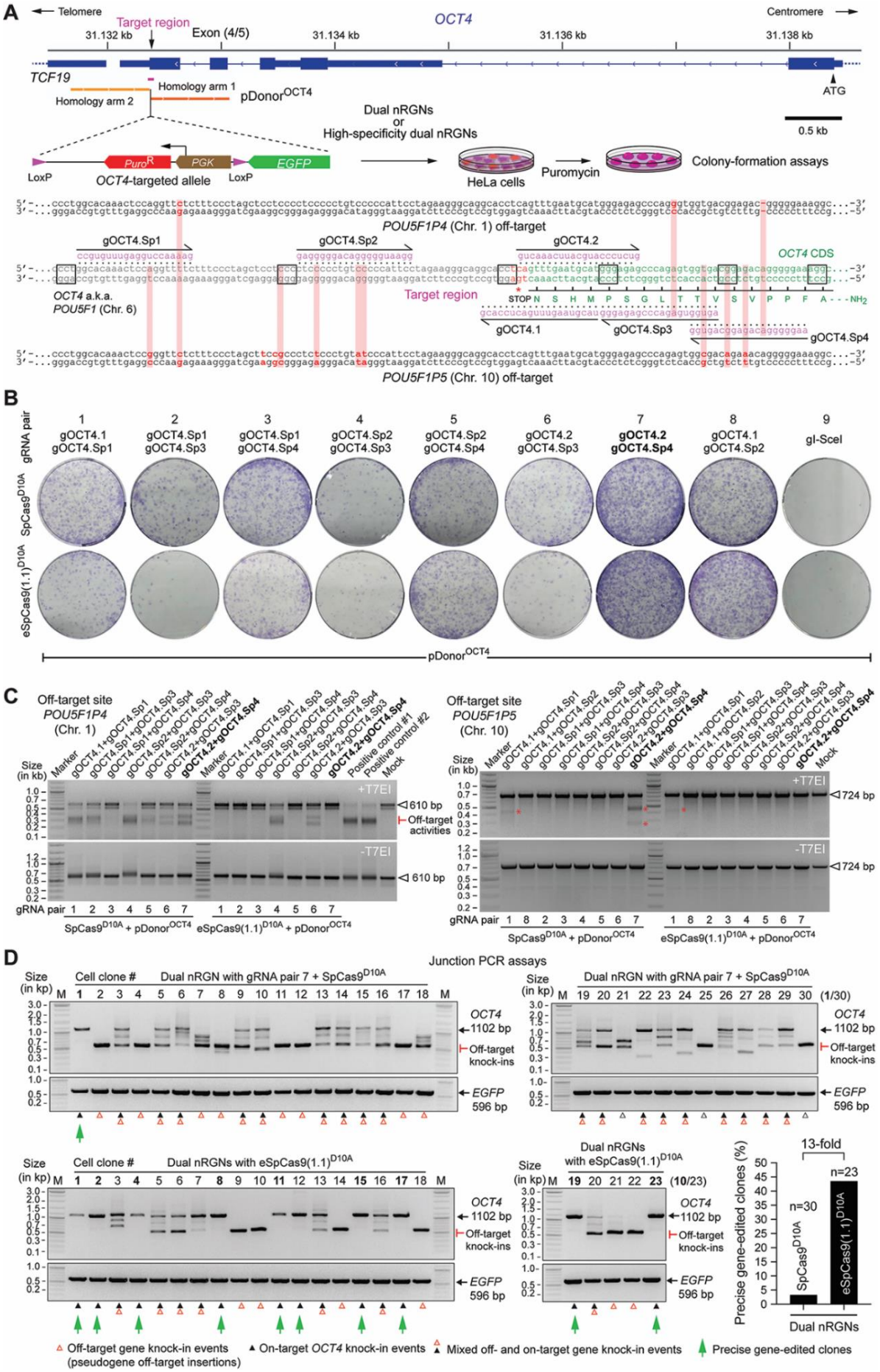

A
Telomere
Centromere
Target region
Exon (4/5)
OCT4
31.132 kb
31.134 kb
31.136 kb
31.138 kb
TCF19
ATG
Homology arm 1
Homology arm 2
pDonorOCT4
0.5 kb
Dual nRGNs
or
High-specificity dual nRGNs
LoxP
PuroR
PGK
EGFP
LoxP
OCT4-targeted allele
HeLa cells
Puromycin
Colony-formation assays
POU5F1P4 (Chr. 1) off-target
gOCT4.Sp1
gOCT4.Sp2
gOCT4.2
OCT4 CDS
OCT4 a.k.a.
POU5F1 (Chr. 6)
Target region
STOP N S H M P S G L T T V S V P P F A - - - NH2
gOCT4.1
gOCT4.Sp3
gOCT4.Sp4
POU5F1P5 (Chr. 10) off-target
B
gRNA pair
SpCas9D10A
eSpCas9(1.1)D10A
1 gOCT4.1 gOCT4.Sp1
2 gOCT4.Sp1 gOCT4.Sp3
3 gOCT4.Sp1 gOCT4.Sp4
4 gOCT4.Sp2 gOCT4.Sp3
5 gOCT4.Sp2 gOCT4.Sp4
6 gOCT4.2 gOCT4.Sp3
7 gOCT4.2 gOCT4.Sp4
8 gOCT4.1 gOCT4.Sp2
9 gI-SceI
pDonorOCT4
C
Off-target site
POU5F1P4
(Chr. 1)
Off-target site
POU5F1P5
(Chr. 10)
Size (in kb)
Marker
gOCT4.1+gOCT4.Sp1
gOCT4.1+gOCT4.Sp2
gOCT4.1+gOCT4.Sp3
gOCT4.1+gOCT4.Sp4
gOCT4.Sp1+gOCT4.Sp2
gOCT4.Sp1+gOCT4.Sp3
gOCT4.Sp1+gOCT4.Sp4
gOCT4.Sp2+gOCT4.Sp3
gOCT4.Sp2+gOCT4.Sp4
gOCT4.2+gOCT4.Sp3
gOCT4.2+gOCT4.Sp4
Positive control #1
Positive control #2
Mock
+T7EI
-T7EI
610 bp
Off-target activities
724 bp
gRNA pair
SpCas9D10A + pDonorOCT4
eSpCas9(1.1)D10A + pDonorOCT4
D
Junction PCR assays
Size (in kp)
Cell clone #
Dual nRGN with gRNA pair 7 + SpCas9D10A
(1/30)
Dual nRGNs with eSpCas9(1.1)D10A
(10/23)
OCT4
1102 bp
Off-target knock-ins
EGFP
596 bp
Precise gene-edited clones (%)
13-fold
n=30
n=23
SpCas9D10A
eSpCas9(1.1)D10A
Dual nRGNs
Off-target gene knock-in events (pseudogene off-target insertions)
On-target OCT4 knock-in events
Mixed off- and on-target gene knock-in events
Precise gene-edited clones

Figure 9. Homology-directed gene targeting of genomic sites sharing high sequence identity with off-target sequences using conventional or high-specificity complexes. (**A**) *OCT4* gene targeting set-up. The *OCT4* target region is presented in relation to similar sequences in *OCT4* pseudogenes *POU5F1P4* and *POU5F1P5* located at chromosomes 1 and 10, respectively. HeLa cells were transfected with pDonor^{OCT4} and plasmids encoding dual nRGNs containing SpCas9 or high-specificity dual nRGNs harboring eSpCas9(1.1)^{D10A}. Donor construct pDonor^{OCT4} is designed to knock-in into *OCT4* the EGFP coding sequence together with a floxed marker gene that confers resistance to puromycin in colony-formation assays. PAM and gRNA sequences are boxed and magenta colored, respectively. DNA-gRNA mismatches are highlighted by vertical red bars. (**B**) Colony-formation assays on HeLa cells. HeLa cells genetically modified through the delivery of the indicated gene-editing tools are scored after puromycin selection and Giemsa staining. (**C**) Detection of dual nRGN off-target activities. T7EI-based genotyping assays were performed on DNA from puromycin-resistant HeLa cell populations initially exposed to pDonor^{OCT4} and the indicated dual nRGN elements. T7EI-specific products diagnostic for mutant *POU5F1P4* and *POU5F1P5* loci generated by the installation of indels after NHEJ-mediated DSB repair, are labelled as "Off-target activities" and asterisks, respectively. Products representing intact loci are instead marked by open arrowheads. (**D**) Characterization of HDR-mediated *OCT4* gene editing specificity achieved by dual nRGNs containing SpCas9^{D10A} or eSpCas9(1.1)^{D10A}. Junction PCR analysis on genomic DNA from puromycin-resistant HeLa cell clones from cultures treated with pDonor^{OCT4}, SpCas9^{D10A}, gOCT4.2 and gOCT4.Sp4 (n=30) or with pDonor^{OCT4}, eSpCas9(1.1)^{D10A}, gOCT4.2 and gOCT4.Sp4 (n=23). For details see **Supplementary Figure S12**. Lanes M, GeneRuler DNA Ladder Mix molecular weight marker.

We proceeded by performing gene knock-in experiments targeting active *OCT4* alleles in iPSCs using pDonor^{OCT4} and gRNA pair members gOCT4.2 and gOCT4.Sp4. The latter gRNA forms a bulge at *POU5F1P4* and displays three mismatches to *POU5F1P5* (**Figure 10A**). The coupling of this gRNA pair to SpCas9^{D10A} or eSpCas9(1.1)^{D10A} yielded high and similar levels of genetically modified HeLa cells (**Figure 9B**). In the OCT4 gene targeting experiments in iPSCs, next to dual nRGNs, we extended the testing to RGNs with SpCas9 or eSpCas9(1.1). The highest numbers of puromycin-resistant iPSCs labeled with the pluripotency marker alkaline phosphatase (AP) were observed in cultures initially exposed to dual nRGNs harboring eSpCas9(1.1)^{D10A} (Hi-Si dual nRGN; **Figure 10B** and **C**). Importantly, off-target DSBs in puromycin-resistance iPSCs subjected to these high-specificity dual nRGNs were detected neither at *POU5F1P4* nor *POU5F1P5* (Hi-Si dual nRGN; **Figure 10D**). In contrast, robust off-target DSB activities at *POU5F1P4* were detected in puromycin-resistant iPSCs subjected to dual nRGNs containing SpCas9^{D10A} (Dual nRGN; **Figure 10D**). In HeLa cells, off-target cleavage provoked by these conventional dual nRGNs was readily detected at *POU5F1P5* as well (**Figure 9C**), possibly reflecting the higher initial transfection efficiencies achieved in these cells.

As expected, RGN complex SpCas9:gOCT4.2 (RGN.1), by presenting complementarity to pseudogene sequences, cleaved *POU5F1P4* and *POU5F1P5* (**Figure 10D**). Notably, despite having the same gRNA as SpCas9:gOCT4.2, off-target cleavage was not detected with eSpCas9(1.1):gOCT4.2 (Hi-Si RGN.1). This result is consistent with the fact that gOCT4.2 has an extended spacer and a 5' non-hybridizing guanine, features previously implicated in eSpCas9(1.1) hindrance here (**Figure 1C**) and elsewhere (24,36-38). Moreover, the highest numbers of AP$^+$ iPSC colonies obtained by using high-specificity dual nRGNs further support our earlier finding that hindrance of eSpCas9(1.1)-mediated DSB formation by non-canonical gRNAs (**Figure 2**, **Supplementary Figures S1** and **S3**) can be overcome, now in a gene knock-in setting, by converting this nuclease into a nickase and placing it in a dual nRGN context (**Figure 10B** and **C**).

Taken together, our results suggest that incorporating eSpCas9(1.1)^{D10A} in dual nRGNs offers the possibility for enhancing the frequencies and specificities of gene knockouts and gene knock-ins, while retaining the broad genomic coverage of dual nRGN designs resulting from their compatibility with wide spacing between nRGNs as well as non-canonical gRNAs. Concerning the latter aspect, as aforesaid, it is possible that non-canonical gRNAs mostly affect the RuvC domain of eSpCas9(1.1) which is rendered dispensable in dual nRGNs with eSpCas9(1.1)^{D10A} (**Figure 2**, **Supplementary Figures S1** and **S3**).

To compare the frequencies of properly targeted *OCT4* alleles in iPSCs genetically modified through RGNs or dual nRGNs with standard or high-specificity enzymes, we exploited the genetic readout system built in pDonor^{OCT4}. In this system, Cre-mediated assembly of a traceable OCT4::EGFP fusion product reports targeted iPSCs in puromycin-resistance populations (**Figure 11A**). Notably, EGFP-directed flow cytometry detected *OCT4*-targeted iPSCs at levels substantially above background exclusively in cell populations genetically modified by standard and high-specificity dual nRGNs (**Figure 11B**). Finally, EGFP and OCT4 confocal microscopy analyses confirmed accurate tagging of the endogenous OCT4 protein in these iPSC populations (**Figure 11C**), which were subsequently capable of differentiating into cells representing the three embryonic germ layers (**Figure 11D**).

Taken together, these data demonstrate that gene-editing involving homologous recombination between pDonor^{OCT4} and *OCT4* was best achieved by using high-specificity dual nRGNs based on eSpCas9(1.1)^{D10A}. In fact, these dual nRGNs outperformed conventional and high-specificity RGNs as well as conventional dual nRGNs in terms of avoiding off-target cleavage at highly similar pseudogene sequences (**Figures 9C** and **10D**) and, at the same time, yielding precise gene knock-ins (**Figures 9D** and **11B**).

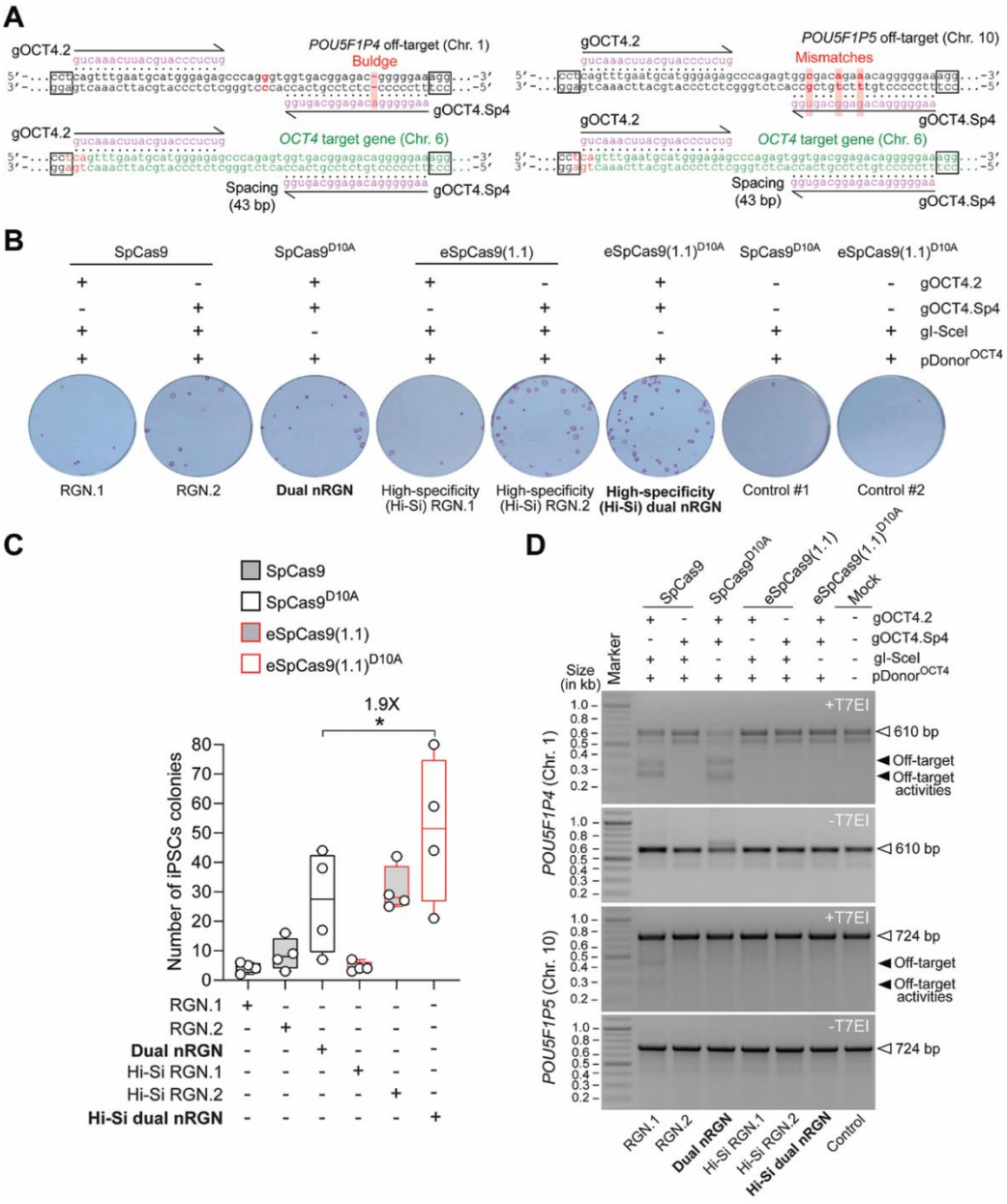


Figure 10. Homology-directed gene targeting in iPSCs at *OCT4* sequences highly similar to off-target sites using conventional or high-specificity complexes. (**A**) RGN and dual nRGN target sites and pseudogene off-target sequences. The *OCT4* sequence (green) is depicted next to similar sequences in *OCT4* pseudogenes *POU5F1P4* and *POU5F1P5* (black) located at chromosomes 1 and 10, respectively. PAM and gRNA sequences are boxed and magenta colored, respectively. DNA-gRNA mismatches and a gRNA buldge are highlighted by vertical red bars. (**B**) Colony-formation assays on iPSCs. iPSCs genetically modified through the transfer of the indicated gene-editing reagents are identified after puromycin selection and staining for the pluripotency marker alkaline phosphatase. (**C**) Quantification of genetically modified iPSCs. The numbers of alkaline phosphatase-positive iPSC colonies resulting from four independent biological replicates are presented in box plots with minimum and maximum. Significance between the indicated datasets was calculated by two-tailed Student's *t* tests; *0.01 < *P* < 0.05 (**D**) Detection of RGN and dual nRGN off-target activities. T7EI-based genotyping assays were carried out on DNA from puromycin-resistant iPSC populations initially subjected to pDonor^{OCT4} and the indicated RGN and dual nRGN components. T7EI-specific species diagnostic for mutant *POU5F1P4* and *POU5F1P5* loci generated by the induction of indels after NHEJ-mediated DSB repair, are marked by solid arrowheads. Products corresponding to intact loci are instead marked by open arrowheads. Marker, GeneRuler DNA Ladder Mix molecular weight marker.

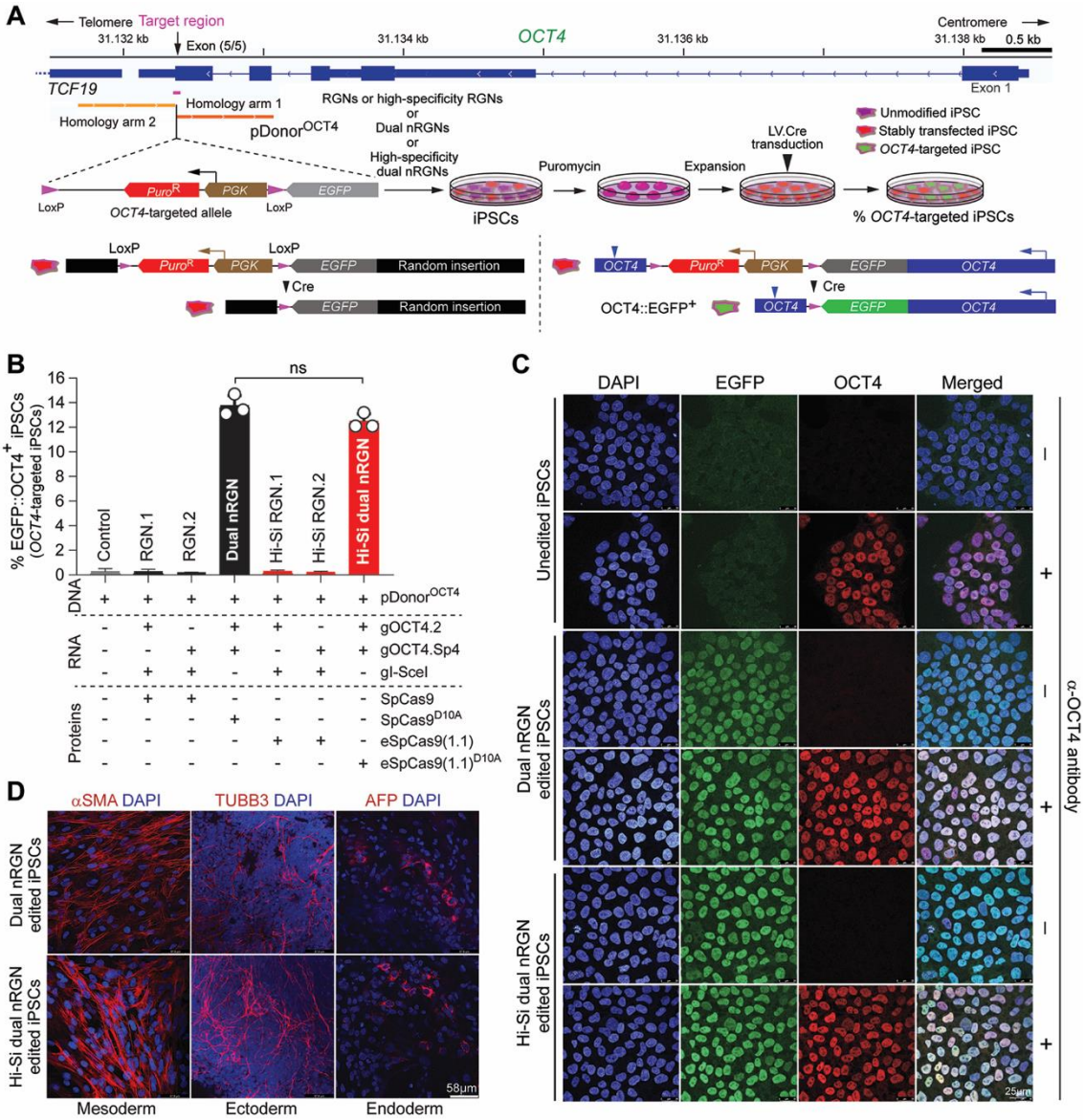


Figure 11. Quantification and characterization of *OCT4* targeted iPSCs by standard versus high-specificity RGNs and dual nRGNs. (**A**) Experimental set-up and genetic assay for detecting *OCT4* gene targeting events. iPSCs were transfected with pDonor^{OCT4} and constructs expressing RGNs containing SpCas9 or eSpCas9(1.1) or high-specificity dual nRGNs harbouring SpCas9^{D10A} or eSpCas9(1.1)^{D10A}. pDonor^{OCT4} knock-ins into *OCT4* the *EGFP* coding sequence and a floxed marker gene conferring puromycin resistance. Functional genetic assays, The Cre-mediated selectable marker removal and OCT4::EGFP fusion product assembly reports precisely targeted iPSCs. Stable OCT4::EGFP expression arises from *OCT4* transcription initiation and termination regulatory elements. (**B**) Quantification of OCT4 targeted iPSCs. The frequencies of OCT4::EGFP^{+} iPSCs in puromycin resistant populations were determined by EGFP-directed flow cytometry. Data are shown as mean ± S.D. of three independent biological replicates. Significance between the indicated datasets was calculated through two-tailed Student's *t* test with $P \geq 0.05$ considered non-significant (ns). (**C**) Confocal fluorescence microscopy analysis of *OCT4* edited iPSCs. OCT4::EGFP^{+} iPSCs edited by dual nRGNs with SpCas9^{D10A} or by high-specificity dual nRGNs with eSpCas9(1.1)^{D10A}, were subjected to indirect and direct fluorescence microscopies for detecting OCT4 and EGFP, respectively. Nuclei were identified by DAPI staining. Parental, unedited, iPSCs served as negative controls. Unedited and edited iPSC populations that were not incubated with the OCT4-specific primary antibody provided for staining controls. (**D**) Testing the multilineage differentiation potential of *OCT4* edited iPSCs. OCT4::EGFP^{+} iPSCs edited by dual nRGNs with SpCas9^{D10A} or by high-specificity dual nRGNs with eSpCas9(1.1)^{D10A} were induced to differentiate into cell lineages corresponding to the three embryonic germ layers, i.e., mesoderm, ectoderm and endoderm. Markers for each of these germ layers are indicated. Nuclei were stained with DAPI.

Unbiased genome-wide assessment of specificity profiles of cleaving versus nicking RGNs

Although most SSBs are resolved through conservative DNA repair processes (20,21), they can nonetheless progress to DSBs in instances in which an advancing replication fork hits them and collapses (49). Therefore, unbiased and sensitive methods for detecting genomic changes resulting from SSBs or nicks are warranted for guiding the refinement of precise gene-editing tools and strategies based on nRGNs. Recently, to measure and examine off-target effects induced by nRGNs, we have

adapted the high-throughput genome-wide translocation sequencing (HTGTS) assay by incorporating SaCas9 nuclease and a universal *RAG1*-targeting gRNA (Sa-gRAG1.1) for inducing bait DSBs (**Figure 12A**) (14). As this assay, dubbed orthogonal HTGTS, permits comparing RGN and nRGN off-target profiles as well, herein we investigated side-by-side the genome-wide specificities of SpCas9, eSpCas9(1.1), SpCas9^{D10A} and eSpCas9(1.1)^{D10A}. Thus, after validating that SpCas9 variants are compatible with the *VEGFA*-targeting gRNA gVEGFA (**Supplementary Figure S13A**), previously used in genome-wide DSB detection assays (23), we introduced this gRNA and universal SaCas9:Sa-gRAG1.1 complexes together with each of the test nucleases or test nickases into HEK293T cells (n = 3). As expected, indels at *RAG1* and *VEGFA* were readily detected in cells exposed to SaCas9:Sa-gRAG1.1 and gVEGFA-bound nucleases (**Supplementary Figure S13B**). In contrast, indels were only detected at *RAG1* in cells subjected to SaCas9:Sa-gRAG1.1 and gVEGFA-bound nickases, confirming that nRGNs have a low mutagenic potential (**Supplementary Figure S13B**). The higher on-target effects induced by nucleases over nickases was independently confirmed by orthogonal HTGTS analysis (**Figure 12B** and **C**, **Supplementary Figures S14** and **S15**). Most importantly, this analysis further demonstrated a gradual overall decrease in off-target effects in cells treated with SpCas9, eSpCas9(1.1), SpCas9^{D10A} and eSpCas9(1.1)^{D10A} (**Figure 13A** and **B**). As expected, SpCas9 was more disruptive to the genome than eSpCas9(1.1) (**Figures 12C** and **13B**, **Supplementary Figures S14** and **S15**). Interestingly, a subtle differential off-target site preference for SpCas9 and eSpCas9(1.1) was uncovered within an enriched translocation region at chromosome 11 (**Supplementary Figures S14** and **S16**). In the case of nicking SpCas9^{D10A} and eSpCas9(1.1)^{D10A} enzymes, off-target activities were detected at two chromosome 14 regions, with the latter enzyme presenting a 2.3-fold lower off-target activity index at one of these two genomic regions (**Figure 13B**, lower panel). Taken together, the orthogonal HTGTS data indicate that, amongst the four proteins tested, eSpCas9(1.1)^{D10A} is the least genome-disrupting thus constituting a preferable tool for precise genome editing based on targeted DSB or SSB formation.

DISCUSSION

We report that the enhanced specificity of a representative panel of SpCas9 mutants is transportable to their respective SpCas9^{D10A} variants. Indeed, albeit differing significantly in their sequence-specific and strand-specific nuclease activities, the assembled RNA-guided nickases exhibit specificities that are markedly superior to that of the commonly used SpCas9^{D10A} protein. By using an array of functional screens, we have identified high-specificity nickases that can, when operating as dual nRGNs, outperform their conventional dual nRGN counterparts in terms of target DNA cleaving activities and specificities. Concerning the latter aspect, after selecting Sniper-Cas9^{D10A}, we provide a proof-of-concept for a three-tier precision gene editing strategy based on integrating into the dual nRGN concept (18,19), the truncated gRNA (40) and high-specificity nickase principles. Moreover, high-specificity dual nRGNs containing eSpCas9(1.1)^{D10A} were found to be more versatile than high-specificity RGNs harboring eSpCas9(1.1). In particular, besides retaining the broad genomic space coverage characteristic of dual nRGN designs, dual nRGNs based on eSpCas9(1.1)^{D10A} were compatible with gRNAs containing extended spacers or 5' non-hybridizing guanines. These data indicate that these non-canonical gRNAs mostly hinder the RuvC domain of eSpCas9(1.1), which is functionally absent in dual nRGNs with eSpCas9(1.1)^{D10A}. Importantly, orthogonal HTGTS analyses detected scant off-target activity at the genome-wide level in cells exposed to eSpCas9(1.1)^{D10A} and the promiscuous gRNA gVEGFA (23). Finally, targeted deep sequencing analysis suggests that the choice of nickase variant and gRNA spacing have an impact on the type and uniformity of 'footprints' installed by dual nRGNs.

A broad range of small and large chromosomal edits can be established following NHEJ or HDR of targeted DSBs. These edits include *de novo* translocations for studying cancer (50), genomic deletions and gene knockouts for investigating cis-acting and *trans*-acting elements, and gene knock-ins to modify, repair or tag endogenous genes (1,5,51,52). However, targeting specific loci or allelic variants in diploid cells is challenging, especially when these elements share high sequence identity with regions located elsewhere in the genome. Yet, for the most part, eukaryotic genomes consist of such recurrent multiple-copy regions that include retroelements, amplified gene clusters, gene paralogs and pseudogenes (53). Moreover, knowledge about genetic differences amongst genomes or amongst different alleles or loci in an individual genome, e.g. SNPs and indels, is crucial for complementing correlative genome-wide association studies (GWAS) with causal genotype-phenotype relationships (54,55). Another aspect concerns the fact that, as genome editing expands its reach into therapeutic gene editing, the human

genetic variation is likely to start receiving further attention. Indeed, it has been shown that SNPs and indels can alter the activity and specificity of RGNs in a genotype-dependent manner, including at loci underpinning human disorders (56). Therefore, there is a pressing need to develop genome editing technologies permitting a judicious access to specific chromosomal sequences while averting related off-target sites. To this end, we exploited genomic indels or SNPs and the heightened single base-pair discriminating power of dual nRGNs with eSpCas9(1.1)^{D10A} to selectively target *OCT4* and avoid off-target *OCT4* pseudogene sequences. In contrast, conventional dual nRGNs readily led to disrupted *OCT4* pseudogene loci. The 'tiptoeing' of dual nRGNs over SNPs permitted retrieving iPSCs expressing EGFP-tagged OCT4. Despite the superior sensitivity of dual nRGNs containing eSpCas9(1.1)^{D10A} to single-base pair mismatches, a limitation of the 'tiptoeing' approach is the need to design and test various gRNA pairs per target region as off-target activities were still detected when using eSpCas9(1.1)^{D10A} and certain gRNA pairs.

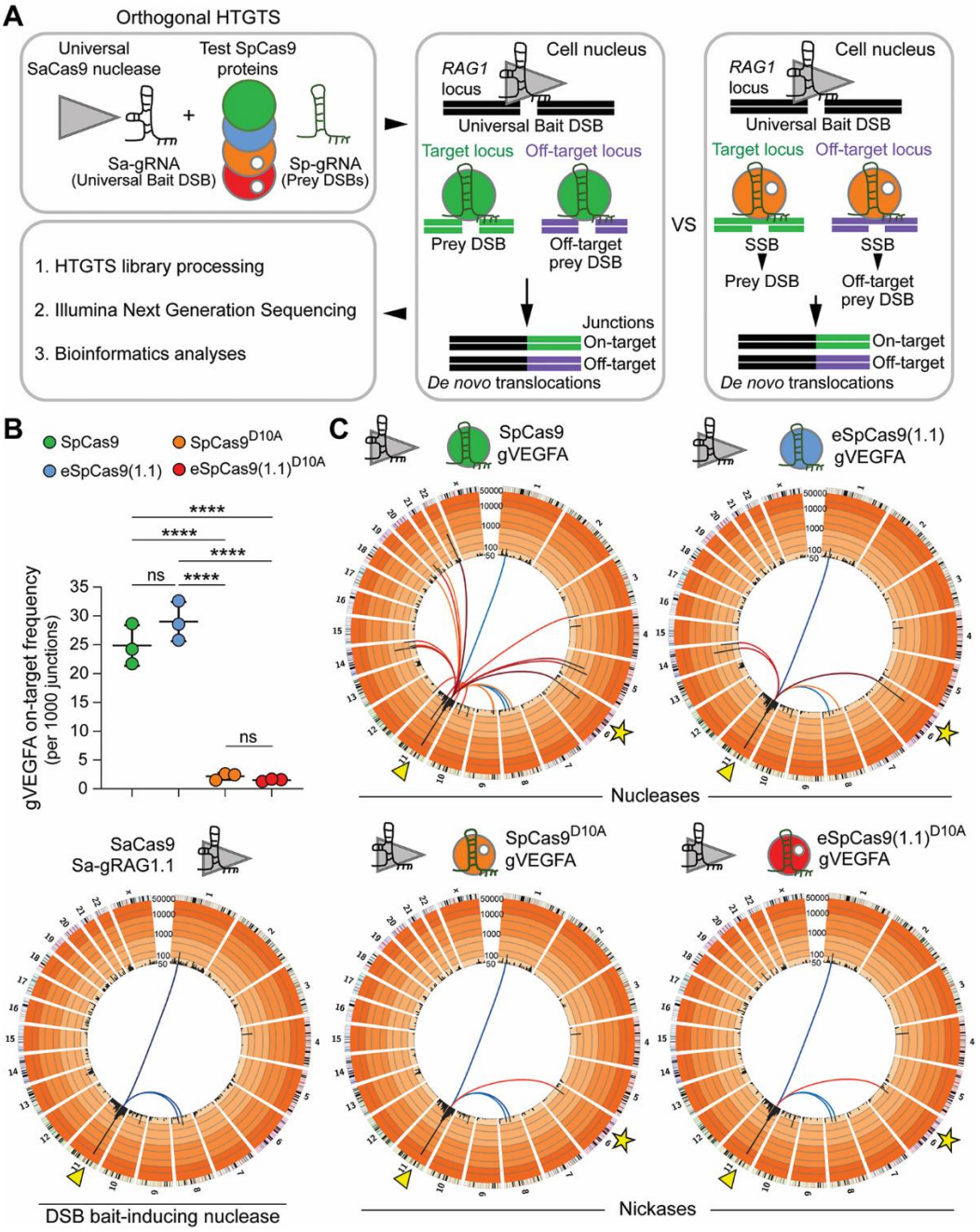


Figure 12. Investigating the specificity of cleaving and nicking RGNs by unbiased genome-wide orthogonal HTGTS analyses. (**A**) Schematics of the orthogonal HTGTS pipeline for genome-wide assessments of off-target effects induced by RGNs versus nRGNs. A universal *S. aureus* cleaving RGN complex (SaCas9:Sa-gRAG1.1) is used to generate bait DSBs at *RAG1*; cleaving and nicking test RGN complexes induce DSBs and SSBs, respectively, at target and off-target loci. Prey DSBs catalyzed

by *S. pyogenes* nucleases and prey DSBs generated from SSBs catalyzed by *S. pyogenes* nickases, are measured through deep sequencing of translocation junctions involving bait and prey chromosomal termini. (**B**) On-target DSB frequencies. Number of translocations to the *VEGFA* target locus per 1000 junctions induced by nucleases SpCas9:gVEGFA and eSpCas9(1.1):gVEGFA or by nickases SpCas9^{D10A}:gVEGFA and eSpCas9(1.1)^{D10A}:gVEGFA. HEK293T cells were transfected with constructs expressing the indicated RGNs and nRGNs (n=3 biological replicates). At 2 days post-transfection, orthogonal HTGTS analyses were carried out on genomic DNA previously screened by target-site genotyping assays (**Supplementary Figure S13B**). ****P<0.0001 one-way ANOVA and Tukey's multiple pairwise-comparisons. (**C**) Cumulative orthogonal HTGTS analyses from three biological replicates. Each library was normalized to 11932 junctions. Arrowheads in Circos plots mark the location of the bait DSB on chromosome 11 induced by the universal *S. aureus* RGN for all sequence read libraries; stars in Circos plots mark the *VEGFA* target site of test *S. pyogenes* RGNs and test *S. pyogenes* nRGNs on chromosome 6. Blue-graded lines connected to the *RAG1* locus indicate bait nuclease-related off-targets; red-graded lines linked to the *RAG1* locus indicate on-target (star) and off-targets resulting from RGNs and nRGNs containing the promiscuous gRNA gVEGFA. Black bars correspond to 5 Mb bins across each chromosome with enrichment levels presented on a custom color-coded log scale by order of magnitude. Hotspots are established when significantly enriched translocation sites are present in at least 2 out of 3 replicates (MACS2; q-value cutoff -10^-10).

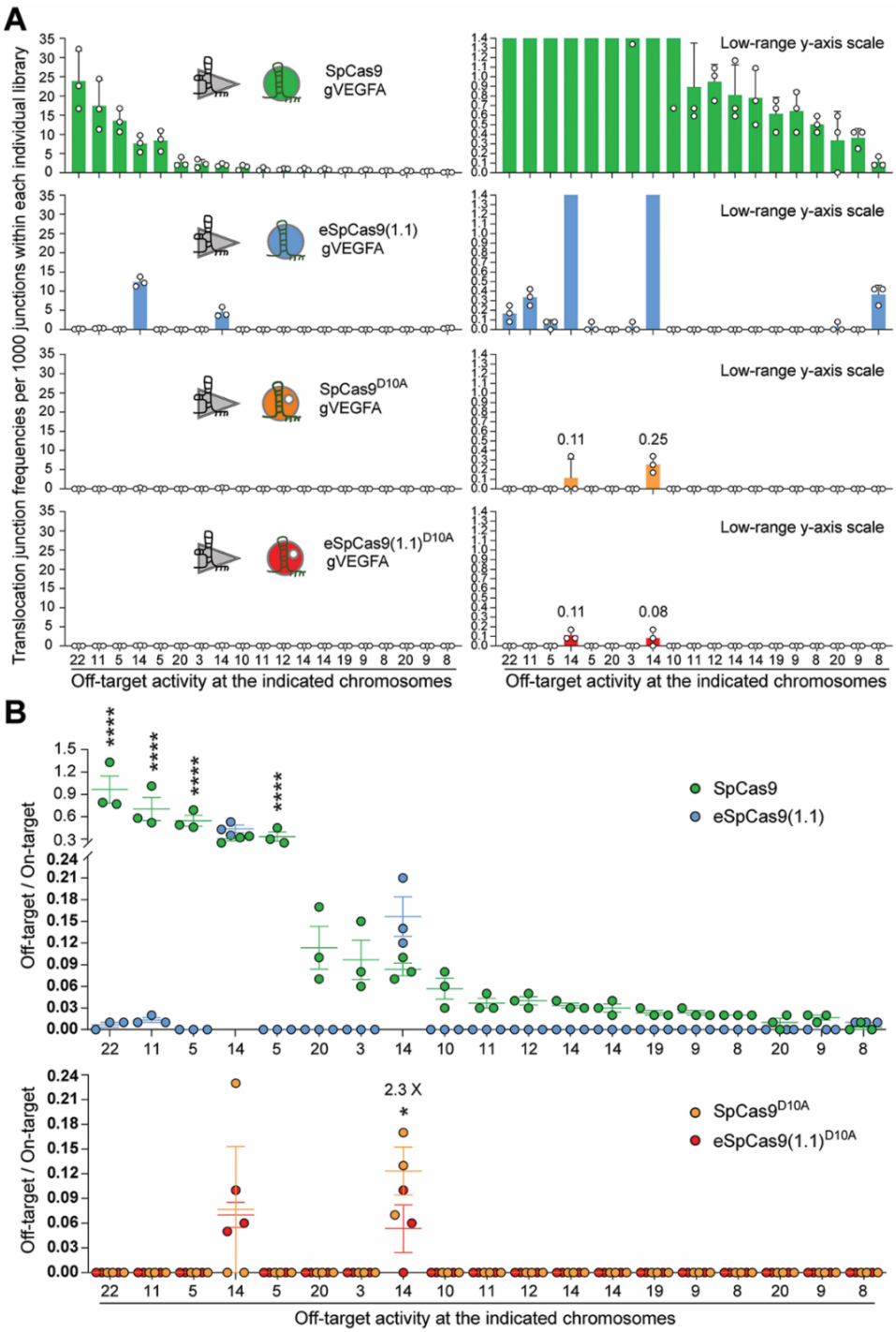


Figure 13. Ranking the off-target sites of RGNs and nRGNs containing a promiscuous gRNA. (**A**) Distribution and frequencies of gVEGFA off-target sites across the human genome. Translocation junction frequencies associated with each of the detected off-target sites plotted with a broad and narrow Y-axis value ranges (left and right panels, respectively). Off-targets were ranked according to their frequencies in sequence read libraries corresponding to SpCas9:gVEGFA complexes. The chromosomes in which each of the off-target sites map are shown. Chromosome coordinates of detected off-target sites and frequencies of translocation junctions per 1,000 junctions within each individual library are specified in **Supplementary Figure S14**. (**B**) Activity indices at the various off-target hotspots. Hotspots are defined as translocation enriched sites found significant in at least 2 out of 3 normalized libraries for each CRISPR complex (MACS2; q-value cutoff -10^-10). Ratios between the number of translocations to an off-target site and the number of translocations to the on-target site at *VEGFA* in libraries normalized to 11932 junctions; asterisks mark statistically significant differences in off-target activity indices in normalized libraries (MACS2; q-value cutoff -10^-10). *P=0.0217; ****P<0.0001 two-way ANOVA and Holm-Sidak's multiple comparison tests. Error bars correspond to mean and SEM from 3 independent biological replicates.

In conclusion, after screening and identifying improved RNA-guided nickases, we demonstrate their utility for expanding precise genomic engineering involving the engagement of the NHEJ and HDR pathways. Recent developments in genome editing entail using nicking Cas9 proteins as such or fused to heterologous DNA-modifying moieties. These genome editing approaches include; (i) HDR-mediated chromosomal insertion of donor DNA spanning from single base-pairs to entire transgenes through nicking of target and donor templates, i.e. *in trans* paired nicking (14,28,57,58), and (ii) donor DNA-free installation of single base-pair transversions through base editing (59–61) and any base-pair substitution or short indel through prime editing (62). The herein investigated high-specificity nickases and gene editing strategies involving the recruitment of either NHEJ or HDR pathways might enrich and complement these emerging technologies directed at seamless and scarless genomic engineering.

DATA AVAILABILITY

All data generated and analysed in this study are included in the article and supplementary files. Additional raw datasets that support the findings of this work are available upon request. The deep sequencing libraries corresponding to the orthogonal HTGTS analyses are deposited in the Gene Expression Omnibus (GEO) repository and are available via accession code GSE153471. The amplicon deep sequencing reads corresponding to dual nicking RGN target-site genotyping analyses are deposited at the NCBI Sequence Read Archive (SRA) database under the BioProject accession PRJNA675830.

SUPPLEMENTARY DATA

Supplementary Data are available online at doi:10.1093/nar/gkaa1236.

ACKNOWLEDGEMENT

The authors thank Chuannan Fan and Steve Cramer (Department of Cell and Chemical Biology, Leiden University Medical Center) for their assistance with the western blot analysis and the research logistics, respectively. The authors also thank Ignazio Maggio (Departments of Pediatrics/Willem-Alexander Kinderziekenhuis and Cell and Chemical Biology, Leiden University Medical Center) and Hailiang Mei (Department of Biomedical Data Sciences, Leiden University Medical Center) for their assistance with the amplicon deep sequencing procedures. Finally, the authors are grateful to the members of their laboratories for the fruitful discussions and support.

FUNDING

Prinses Beatrix Spierfonds (W.OR16-13); Dutch Duchenne Parent Project (17.012); Q.W. holds a Ph.D. fellowship from the China Scholarship Council–Leiden University Joint Scholarship Programme. Funding for open access charge: Prinses Beatrix Spierfonds (W.OR16-13).
Conflict of interest statement. None declared.

REFERENCES

1. Jiang, F. and Doudna, J.A. (2017) CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.*, **46**, 505-529.
2. Anders, C., Niewoehner, O., Duerst, A. and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, **513**, 569-573.
3. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.

4. 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. *et al.* (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, **520**, 186-191.
5. Chandrasegaran, S. and Carroll, D. (2016) Origins of Programmable Nucleases for Genome Engineering. *J. Mol. Biol.*, **428**, 963-989.
6. Cradick, T.J., Fine, E.J., Antico, C.J. and Bao, G. (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.*, **41**, 9584-9592.
7. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822-826.
8. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827-832.
9. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839-843.
10. Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S. and Kim, J.S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.*, **24**, 132-141.
11. Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J. and Bao, G. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.*, **42**, 7473-7485.
12. Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E. and Alt, F.W. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.*, **33**, 179-186.
13. Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., Jin, Z.B., Qu, J. and Gu, F. (2014) Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.*, **4**, 5405.
14. Chen, X., Tasca, F., Wang, Q., Liu, J., Janssen, J.M., Brescia, M.D., Bellin, M., Szuhai, K., Kenrick, J., Frock, R.L. *et al.* (2020) Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking. *Nucleic Acids Res.*, **48**, 974-995.
15. Chen, X. and Goncalves, M. (2018) DNA, RNA, and Protein Tools for Editing the Genetic Information in Human Cells. *iScience*, **6**, 247-263.
16. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.
17. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
18. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.*, **31**, 833-838.
19. 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. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380-1389.
20. Caldecott, K.W. (2014) DNA single-strand break repair. *Exp. Cell Res.*, **329**, 2-8.
21. Vriend, L.E. and Krawczyk, P.M. (2017) Nick-initiated homologous recombination: Protecting the genome, one strand at a time. *DNA Repair (Amst)*, **50**, 1-13.
22. Gopalappa, R., Suresh, B., Ramakrishna, S. and Kim, H.H. (2018) Paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption. *Nucleic Acids Res.*, **46**, e71.
23. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X. and Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84-88.
24. Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.H., Lee, K., Jung, I., Kim, D., Kim, S. *et al.* (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.*, **9**, 3048.

25. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490-495.
26. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F. *et al.* (2018) A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.*, **36**, 265-271.
27. Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z. *et al.* (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556**, 57-63.
28. Chen, X., Janssen, J.M., Liu, J., Maggio, I., t Jong, A.E.J., Mikkers, H.M.M. and Goncalves, M. (2017) In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun.*, **8**, 657.
29. Goncalves, M.A., van der Velde, I., Knaan-Shanzer, S., Valerio, D. and de Vries, A.A. (2004) Stable transduction of large DNA by high-capacity adeno-associated virus/adenovirus hybrid vectors. *Virology*, **321**, 287-296.
30. Chen, X., Rinsma, M., Janssen, J.M., Liu, J., Maggio, I. and Goncalves, M.A. (2016) Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.*, **44**, 6482-6492.
31. Chen, X., Liu, J., Janssen, J.M. and Goncalves, M. (2017) The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Mol. Ther. Nucleic Acids*, **8**, 558-563.
32. Maggio, I., Zittersteijn, H.A., Wang, Q., Liu, J., Janssen, J.M., Ojeda, I.T., van der Maarel, S.M., Lankester, A.C., Hoeben, R.C. and Goncalves, M. (2020) Integrating gene delivery and gene-editing technologies by adenoviral vector transfer of optimized CRISPR-Cas9 components. *Gene Ther.*, **27**, 209-225.
33. Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W. and Frock, R.L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.*, **11**, 853-871.
34. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E. et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.*, **37**, 224-226.
35. Goedhart, J., von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M.A., van Weeren, L., Gadella, T.W., Jr. and Royant, A. (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.*, **3**, 751.
36. Kulcsar, P.I., Talas, A., Huszar, K., Ligeti, Z., Toth, E., Weinhardt, N., Fodor, E. and Welker, E. (2017) Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol.*, **18**, 190.
37. Kim, S., Bae, T., Hwang, J. and Kim, J.S. (2017) Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. *Genome Biol.*, **18**, 218.
38. Zhang, D., Zhang, H., Li, T., Chen, K., Qiu, J.L. and Gao, C. (2017) Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol.*, **18**, 191.
39. Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X. *et al.* (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods*, **11**, 399-402.
40. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. and Joung, J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.*, **32**, 279-284.
41. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861-872.
42. Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. and Benvenisty, N. (2018) Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol.*, **20**, 610-619.
43. Fogarty, N.M.E., McCarthy, A., Snijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D. *et al.* (2017) Genome editing reveals a role for OCT4 in human embryogenesis. *Nature*, **550**, 67-73.
44. Liu, J.C., Lerou, P.H. and Lahav, G. (2014) Stem cells: balancing resistance and sensitivity to DNA damage. *Trends Cell Biol.*, **24**, 268-274.

45. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C. *et al.* (2018) p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.*, **24**, 939-946.
46. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. and Taipale, J. (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.*, **24**, 927-930.
47. Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C. *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.*, **29**, 731-734.
48. Zhu, Z., Verma, N., Gonzalez, F., Shi, Z.D. and Huangfu, D. (2015) A CRISPR/Cas-Mediated Selection-free Knockin Strategy in Human Embryonic Stem Cells. *Stem Cell Reports*, **4**, 1103-1111.
49. Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8241-8246.
50. Brunet, E. and Jasin, M. (2018) Induction of Chromosomal Translocations with CRISPR-Cas9 and Other Nucleases: Understanding the Repair Mechanisms That Give Rise to Translocations. *Adv. Exp. Med. Biol.*, **1044**, 15-25.
51. Maggio, I. and Goncalves, M.A. (2015) Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends. Biotechnol.*, **33**, 280-291.
52. Chen, X. and Goncalves, M.A. (2016) Engineered Viruses as Genome Editing Devices. *Mol. Ther.*, **24**, 447-457.
53. Sedlazeck, F.J., Lee, H., Darby, C.A. and Schatz, M.C. (2018) Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.*, **19**, 329-346.
54. Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J. and Lachance, J. (2018) Genetic disease risks can be misestimated across global populations. *Genome Biol.*, **19**, 179.
55. Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G. and Meyre, D. (2019) Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.*, **20**, 467-484.
56. Lessard, S., Francioli, L., Alfoldi, J., Tardif, J.C., Ellinor, P.T., MacArthur, D.G., Lettre, G., Orkin, S.H. and Canver, M.C. (2017) Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E11257-E11266.
57. Nakajima, K., Zhou, Y., Tomita, A., Hirade, Y., Gurumurthy, C.B. and Nakada, S. (2018) Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res.*, **28**, 223-230.
58. Hyodo, T., Rahman, M.L., Karnan, S., Ito, T., Toyoda, A., Ota, A., Wahiduzzaman, M., Tsuzuki, S., Okada, Y., Hosokawa, Y. *et al.* (2020) Tandem Paired Nicking Promotes Precise Genome Editing with Scarce Interference by p53. *Cell Rep.*, **30**, 1195-1207 e1197.
59. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y. *et al.* (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, **353**.
60. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420-424.
61. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I. and Liu, D.R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*, **551**, 464-471.
62. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram A. *et al*. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149-157.

Chapter 4

Precise homology-directed installation of large genomic edits in human cells with cleaving and nicking high-specificity Cas9 variants

Qian Wang[1], Jin Liu[1], Josephine M. Janssen[1] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands

Nucleic Acids Research, 2023, Vol.51, No.7, ***3465–3484***
doi: 10.1093/nar/gkad165

ABSTRACT

Homology-directed recombination (HDR) between donor constructs and acceptor genomic sequences cleaved by programmable nucleases, permits installing large genomic edits in mammalian cells in a precise fashion. Yet, next to precise gene knock-ins, programmable nucleases yield unintended genomic modifications resulting from non-homologous end-joining processes. Alternatively, in trans paired nicking (ITPN) involving tandem single-strand DNA breaks at target loci and exogenous donor constructs by CRISPR-Cas9 nickases, fosters seamless and scarless genome editing. In the present study, we identified high-specificity CRISPR-Cas9 nucleases capable of outperforming parental CRISPR-Cas9 nucleases in directing genome editing through homologous recombination (HR) and homology-mediated end joining (HMEJ) with donor constructs having regular and 'double-cut' designs, respectively. Additionally, we explored the ITPN principle by demonstrating its compatibility with orthogonal and high-specificity CRISPR-Cas9 nickases and, importantly, report that in human induced pluripotent stem cells (iPSCs), in contrast to high-specificity CRISPR-Cas9 nucleases, neither regular nor high-specificity CRISPR-Cas9 nickases activate P53 signaling, a DNA damage-sensing response linked to the emergence of gene-edited cells with tumor-associated mutations. Finally, experiments in human iPSCs revealed that differently from HR and HMEJ genome editing based on high-specificity CRISPR-Cas9 nucleases, ITPN involving high-specificity CRISPR-Cas9 nickases permits editing allelic sequences associated with essentiality and recurrence in the genome.

INTRODUCTION

Owing to their versatility and potency, RNA-programmable nucleases derived from bacterial CRISPR-Cas9 adaptive immune systems are offering numerous opportunities in basic and applied research, including for the development of genetic therapies (1). Engineered CRISPR-Cas9 nucleases consist of a single guide RNA (gRNA) and a Cas9 enzyme with HNH and RuvC catalytic domains. In the growing set of CRISPR-based genome editing tools, prototypic *Streptococcus pyogenes* Cas9 (SpCas9) and its smaller orthologue *Staphylococcus aureus* Cas9 (SaCas9) nucleases are amongst the most robust (2,3). In cells, these ribonucleoprotein complexes start by engaging short genomic sequences named protospacer adjacent motifs (PAMs) that read NGG and NNGRRT (R = A or G) in the case of SpCas9 and SaCas9, respectively (3,4). Site-specific double-stranded DNA break (DSB) formation follows when, next to a PAM, lies a sequence (protospacer) complementary to the 5′ end of the gRNA (spacer). Specifically, after Cas9-PAM binding and local DNA unwinding, gRNA:DNA hybridization forms a R-loop whose progression from a PAM-proximal to PAM-distal direction eventually overcomes a conformational checkpoint barrier that triggers HNH translocation to the RuvC domain and DNA cleavage (4–6). Modulation of this conformational activation checkpoint by rationally designed or molecularly evolved Cas9 variants achieves heightened DNA mismatch discrimination and hence blunted off-target activities (7–11). As such, these mutant Cas9 enzymes constitute a critical resource for improving genome editing protocols, including those investigated in this study based on the targeted insertion of exogenous (donor) DNA into predefined chromosomal positions (12–14). Indeed, these genome editing approaches are appealing in that they permit introducing genomic modifications spanning from single base-pairs to whole transgene(s) and build on the straightforward designing of RNA-programmable Cas9 nucleases known to have high activities in mammalian cells (3,15–18).

Typically, CRISPR-Cas9 implementation of large genomic edits is accomplished by delivering donor DNA constructs tailored for site-specific DSB repair through ectopic non-homologous end joining (NHEJ) (19,20) or homology-directed repair (HDR) processes (12–14). The latter processes engage donor constructs favoring homologous recombination (HR) (12,13), microhomology-mediated end joining (MMEJ) (21) or, more recently, homology-mediated end joining (HMEJ) (22–24). MMEJ, HMEJ and HR donors have homology tracts ('arms') flanking the exogenous DNA whose sizes span approximately 20–50, 50–900 and 0.5–2.0 kb, respectively. In contrast, NHEJ-prone donors lack sequence homology to target DNA (19,20). In addition, diversely from HR donors, donors prone to NHEJ, MMEJ and HMEJ have a 'double-cut' design in that their targeting modules are surrounded by CRISPR-Cas9 cleaving sites (12–14). This design guarantees exogenous DNA release from construct backbones in cell nuclei, presumably fostering its exonucleolytic processing and target sequence annealing. Importantly, experimental evidence indicates that HR and HMEJ donors yield more precise and directional DNA insertions than their MMEJ and NHEJ counterparts (19,21,22). Additional data further shows that HMEJ donors can lead to higher genome editing frequencies than HR, MMEJ and NHEJ donors in mammalian cells and mouse blastocysts (22–24).

In this work, we start by identifying high-specificity CRISPR-Cas9 nucleases that once combined with donors strictly susceptible to HR or to HMEJ processes, trigger genome editing at levels similar to or higher than those obtained with regular CRISPR-Cas9 nucleases. Additional experiments established that high-specificity CRISPR-Cas9 complexes yield on-target and precise chromosomal insertion of large genetic payloads in human induced pluripotent stem cells (iPSCs). However, as expected, a substantial fraction of target alleles contained small insertions and deletions (indels) due to the prevalence of non-homologous end-joining (NHEJ) pathways over HDR in mammalian cells (25). Besides constituting substrates for mutations and chromosomal rearrangements (26,27), DSBs can lead to haploinsufficiency and cell fitness losses, e.g. when located in exons (28). Moreover, P53-dependent cell cycle arrest and apoptosis induced by CRISPR-Cas9-derived DSBs limits the efficacy of genome editing in stem cells (29,30), confounds genetic screens and, critically, creates selective pressure for the emergence of P53 and KRAS mutations which raises safety risks in stem cell therapies (31–33).

Cas9 proteins with either one of their nuclease domains disabled act as sequence-specific and strand-specific nucleases (nickases). Cas9 nickases are particularly appealing genome editing tools in that single-stranded DNA breaks (SSBs), or nicks, as such are not engaged by mutagenic end-joining DNA repair processes. Moreover, although chromosomal nicks constitute poor HDR stimuli, earlier research in our laboratory uncovered that tandem nicking at endogenous target sites and donor DNA constructs by native or engineered nickases elicits HDR-mediated genome editing (24,34). Examples concerning the application of such in trans paired nicking (ITPN) principles include mutation repair or installation (35–38), allele-specific gene editing (39,40), one-step biallelic gene editing (24,41), and one-step multiplexing gene knock-in or tagging (24,41) in various cell types, e.g. iPSCs, recessive dystrophic epidermolysis bullosa keratinocytes and organoids with regular or cancer traits (24,36,39,41).

Although nicks are mostly resolved in a conservative manner, they can nonetheless lead to DSBs if a replication fork advances through them and collapses (42). It is also known that the extent of baseline indel formation by Cas9 nickases vary in a locus sequence-dependent manner (43). Moreover, in previous studies from our laboratory, unbiased high-throughput genome-wide translocation sequencing (HTGTS) revelated that, albeit at low frequencies, SpCas9^{D10A}:gRNA complexes can trigger translocations involving gRNA off-target sites and that using high-specificity SpCas9^{D10A}:gRNA complexes can further reduce these unwanted genomic effects (28,44). Thus, towards expanding the application of ITPN genome editing and further minimizing nickase-derived DSBs at off-target sequences, we proceeded by investigating its compatibility with SaCas9 nickases and a set of high-specificity SpCas9 nickases. Finally, we found that in contrast to genome editing based on regular and high-specificity CRISPR-Cas9 nucleases, neither regular nor high-specificity CRISPR-Cas9 nickases provokes the P53-dependent DNA damage response (DDR) in human iPSCs.

MATERIAL AND METHODS

Cells

Human cervix carcinoma HeLa cells (American Type Culture Collection) and human embryonic kidney HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific; Cat. No.: 41966029) supplemented with 5% fetal bovine serum (FBS). The generation, characterization and culture conditions of the human iPSCs used in this study (LUMC0020iCTRL) were detailed elsewhere (24,45). In brief, the iPSCs were kept in Essential 8 Medium (E8; Thermo Fisher Scientific; Cat. No.: A1517001) supplemented with 25 U ml^{-1} penicillin and 25 μg ml^{-1} of streptomycin (Thermo Fisher Scientific; Cat. No.: 15140122). Vitronectin Recombinant Human Protein (VTN-N; Thermo Fisher Scientific; Cat. No.: A14700) was applied for coating all the vessels used for iPSC culturing. The different cell types were tested for the absence of mycoplasma contamination and were cultured at 37ºC in humidified-air atmospheres with 5% CO_2 (human iPSCs) or 10% CO_2 (HeLa and HEK293T cells).

Recombinant DNA

Standard recombinant DNA techniques were applied for the generation of the various expression plasmids. The assembly of isogenic expression constructs encoding the different SpCas9 nucleases and SpCas9^{D10A} nickases was described previously (44). Additionally, except for BA32_pU.CAG.SaCas9^{N580A}, the generation of expression constructs encoding *S. aureus* SaCas9

nuclease and SaCas9^{D10A} nickase, was also detailed elsewhere (44). The annotated maps and nucleotide sequences of BA32_pU.CAG.SaCas9^{N580A}, BB43_pmC.Donor^{R5}, BB44_pmC.Donor$^{R5.TS}$, AT13_pE.Donor^{S1}, AA63_pE.Donor$^{S1.TS}$, BA02_pE.Donor^{CLYBL}, AZ64_pE.Donor$^{CLYBL.TS}$, AD60_pEP.Donor^{CLYBL} and AD59_pEP.Donor$^{CLYBL.TS}$ are available in pages 1–27 of the **Supplementary Information**. Detailed information about the *AAVS1*-targeting donor plasmids AX44_pS.Donor^{S1} (#100289), AX53_pS.Donor$^{S1.TS}$ (#100290), AV11_pDonor.EP^{S1} (#100296) and AV09_ pDonor.EP$^{S1.TS}$ (#100297), is available in an earlier work from our laboratory (24), and through the Addgene repository. Likewise for accessing information about AY22_pgRNA^{R5} (#100294) and AS11_gRNA^{S1} (#41818), encoding *CCR5*-specific and *AAVS1*-specific gRNAs, respectively. The generation of *OCT4*-targeting gRNA and donor constructs was described previously (28). Finally, specifics about the gRNA negative control constructs gRNA Cloning Vector (#41824) (18) and BPK2660 (#70709) (46), herein named, gRNA^{Empty} and Sa-gRNA^{Empty}, respectively, can be equally obtained from Addgene. The target sequences of the *S. pyogenes* gRNAs and *S. aureus* Sa-gRNAs used in this work are indicated in **Supplementary Table S1**.

DNA transfections

HeLa cells were transfected with the aid of 1 mg ml^{-1} 25 kDa linear polyethyleneimine (PEI, Polysciences) solution (pH 7.4) following the protocol described previously (44). The transfections of iPSCs were done by using the Lipofectamine Stem Transfection Reagent (Thermo Fisher Scientific; Cat. No.: STEM00003) according to the manufacturer's protocol. The cell numbers and the compositions of different transfection reactions are specified in **Supplementary Tables S2-S22**.

Target-site genotyping

Genotyping assays assessing HDR-mediated knock-ins were performed through restriction fragment length analyses (RFLA) and junction PCR. RFLA assays were initiated by amplifying amplicons spanning the target sequences with the primers and PCR cycling conditions indicated in **Supplementary Tables S23** and **S24**, respectively. Subsequently, 10 μl of the resulting PCR mixtures were incubated with 1 μl (10 U) of the restriction enzyme HindIII (Thermo Fisher Scientific; Cat. No.: ER0501) overnight at 37°C and were then analysed by agarose gel electrophoresis with the aid of a Gel-Doc XR+ system and the ImageLab 6.0.1 software (both from Bio-Rad). Undigested samples served as negative controls. The primer sequences and PCR cycling conditions used for junction PCR analyses are listed in **Supplementary Tables S25** and **S26**, respectively.

Flow cytometry

Nuclease- and nickase-trigged genome editing frequencies were determined by using a BD LSR II flow cytometer (BD Biosciences). Briefly, cells were harvested and resuspended in PBS supplemented with 0.5% bovine serum albumin (BSA) and 2 mM EDTA (pH 8.0). Parental non-transfected cells served as negative controls to establish the thresholds for background fluorescence. At least 10,000 viable single cells were acquired per sample. Data were analyzed with the aid of the FlowJo software (Tree Star; version 10.5.0). The genome editing frequencies were normalized to the initial transfection efficiencies as determined at 3 days post-transfection by using reporter-directed flow cytometry (**Supplementary Figure S1**).

Amplicon deep sequencing

Mutagenic loads in cells edited through canonical HR versus ITPN were assessed using amplicon deep sequencing following the protocol detailed previously (44). The primers, cycling parameters and PCR mixtures used for the preparation of gene-specific and barcoded amplicons are indicated in **Supplementary Tables S27–S30**. Finally, amplicons were pooled in equal molar ratios and were subjected to next-generation Illumina MiSeq deep sequencing for obtaining 100,000 paired-end reads. The frequencies of on-target and off-target genomic indels were quantified with the aid of the CRISPResso2 software (47) after demultiplexing and adapter trimming of the paired-end MiSeq raw reads (R1 and R2 fastq files) with Cutadapt version 2.10 (48). The scripts applied for the CRISPResso2 analyses are available in the **Supplementary Information**.

Off-target donor DNA insertion analysis

HeLa cells were transfected with constructs designed for HR, HMEJ or ITPN at *AAVS1* following the scheme specified in **Supplementary Table S20**. At 10 days post-transfection, the cells were exposed

to puromycin (Thermo Fisher Scientific; Cat. No.: A1113803) at a final concentration of 1 μg ml^{-1} after which puromycin-resistant cell populations were harvested for genomic DNA extraction by using the DNeasy® Blood & Tissue Kit according to the manufacturer's instructions (QIAGEN; Cat. No.: 69506). Donor DNA insertions at off-target *CPNE5* and at target *AAVS1* sequences were captured by junction PCR assays with the aid of Platinum™ SuperFi II DNA Polymerase (Thermo Fisher Scientific; Cat. No.: 12361010). Amplicons specific for *EGFP* served as internal controls. The PCR primers and cycling conditions used in these junction PCR assays are listed in **Supplementary Tables S31** and **S32**, respectively. Afterwards, 10 μl of the *CPNE5* amplicons were incubated overnight at 37ºC with 10 U of the restriction enzymes EcoRI (Thermo Fisher Scientific; Cat. No.: ER0271) and PstI (Thermo Fisher Scientific; Cat. No.: ER0615) and were then analyzed by agarose gel electrophoresis with the aid of a Gel-Doc XR+ system and the ImageLab 6.0.1 software (both from Bio-Rad). In addition, indel formation at genomic target sequences was probed in cells edited through canonical HR, HMEJ and ITPN. To this end, the *AAVS1* target region was amplified using the PCR primers and cycling conditions indicated in **Supplementary Tables S23-S24**, and the resulting PCR products were then subjected to genotyping assays based on the mismatch-sensing T7 endonuclease I (T7EI). In brief, T7EI assays were initiated by subjecting *AAVS1* amplicons to the cycling parameters indicated in **Supplementary Table S33** and, subsequently, 10-μl samples were treated with 0.5 μl (5 U) of the T7EI enzyme (New England Biolabs; Cat. No.: M0302) for 15 min at 37°C. T7EI-digested and undigested DNA was analyzed after agarose gel electrophoresis by using a Gel-Doc XR+ system and the ImageLab 4.1 software (both from Bio-Rad). Finally, Sanger sequencing of *AAVS1* amplicons followed by Tracking of Indels by Decomposition (TIDE) (49) was equally applied to probe indel formation in puromycin-resistant HeLa cell populations edited through ITPN.

IncuCyte cell proliferation assay

iPSCs were seeded at a density of 2×10^3 cells per well of 96-well plates coated with VTN-N. After approximately 16 h, the cells were exposed to 10 μM Nutlin-3a (Cayman Chemical; Cat. No: 675576-98-4) or to the vehicle DMSO for three days. Cell proliferation activity was monitored in the IncuCyte live-cell imaging system and real-time analyzed by the IncuCyte software (Essen BioScience).

Cell viability assay

The colorimetric MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) assay was carried out for assessing iPSC viability upon Nutlin-3a treatment. In brief, iPSCs were seeded at a density of 2×10^3 cells per well of 96-well plates coated with VTN-N. The next day, the cells were exposed to 2 μM Nutlin-3a, 10 μM Nutlin-3a or to DMSO vehicle for 6, 24, 48 and 72 h. Mock-treated iPSCs served as negative controls. At each of the indicated timepoints, 20 μl of MTS solution (Promega; Cat. No.: G3581) were directly added to each sample and the resulting mixtures were then incubated for 1 h at 37°C. The absorbance at OD_{490} nm was measured with the aid of a multimode plate reader (PerkinElmer VICTOR™ X3).

Apoptosis analysis

The frequency of apoptotic iPSCs was quantified by using an eBioscience™ Annexin V Apoptosis Detection Kit FITC (Thermo Fisher Scientific; Cat. No.: 88–8005-72) following the manufacturer's recommendations. In brief, iPSCs were plated at a density of 1×10^5 cells per well of 12-well plates coated with VTN-N. After a 2-day incubation period, the cells were treated with 10 μM Nutlin-3a for 4 h, 6 h, and 8 h. Cells exposed to the protein kinase inhibitor Staurosporine (Cell Signaling Technology; Cat. No: 9953S) or to DMSO vehicle served as positive and negative controls for apoptosis, respectively. At the indicated timepoints, the iPSCs were harvested and resuspended in 1× Binding Buffer. Subsequently, each cell suspension was incubated for 10 min at room temperature with 5 μl of Annexin V conjugated to the FITC fluorochrome. After washing twice with 1× Binding Buffer, the cells were resuspended in 200 μl of 1× Binding Buffer containing 10 μl of 20 μg ml^{-1} propidium iodide (PI). Finally, the frequency of apoptotic iPSCs was determined by using a BD LSR II flow cytometer (BD Biosciences) with the acquired data being analysed with the aid of the FlowJo software (Tree Star; version 10.5.0).

Real-time quantitative PCR (RT-qPCR)

RT-qPCR was applied for assessing the activation of the P53-dependent DDR. Total RNA was extracted by using the NucleoSpin RNA Kit according to the manufacturer's instructions (Macherey Nagel; Cat. No.: 740955). Equal amounts of isolated RNA quantified with a NanoDrop apparatus were reverse

transcribed by using the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific; Cat. No.: K1691). In brief, 500–1000 ng of RNA were incubated with 0.5 μl of 100 μM random hexamer primers and 0.5 μl of 100 μM Oligo(dT)$_{18}$ primers in 12-μl reaction volumes at 65°C for 5 min followed by 2-min incubations at 4°C. Subsequently, 1 μl of 20 U μl^{-1} RiboLock RNase Inhibitor, 1 μl of 200 U μl^{-1} RevertAid H Minus M-MuLV Reverse Transcriptase, 2 μl of 10 mM dNTP Mix and 4 μl of 5× Reaction Buffer, were directly added to each sample and the resulting mixtures were then incubated for 5 min at 25°C followed by 1 h at 42°C. Next, after deactivating the reverse transcriptase by heating at 70°C for 5 min, 1 μl of the synthesized cDNA samples was subjected to qPCR using the iQ™ SYBR® Green Supermix (Bio-Rad; Cat. No.: L010171C) for determining the expression of TP53 and of the canonical P53-responsive genes *P21*, *FAS*, *PUMA* and *MDM2* as well as of the P53 non-responsive gene *HPRT1*. Housekeeping *GAPDH* transcripts were targeted to serve as references for expression normalization. The specificity of each primer pair was predicted by in silico BLAST screens and then validated by assessing the melting profiles. Information on target sequences, qPCR primers, mixture components and reaction conditions are indicated in **Supplementary Tables S34** and **S35**. The CFX Connect Real-Time PCR Detection System (Bio-Rad) was applied for the detection of signal outputs and the relative expression levels were calculated through the $2^{-\Delta\Delta Ct}$ method with the aid of the Bio-Rad CFX Manager software (version 3.1). The GraphPad Prism software (version 9.3.1) was applied for the statistical analyses of the resulting RT-qPCR datasets.

Western blotting

Laemmli buffer consisting of 8.0% glycerol, 3% sodium dodecyl sulphate (SDS) and 200 mM Tris–HCl (pH 6.8) was applied for lysing human iPSCs and HEK293T cells. Afterwards, equal amounts of protein were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto 0.45 μm polyvinylidene difluoride (PVDF) membranes (Merck Millipore; Cat. No.: IPVH00010). After 1 h blocking at room temperature in Tris-buffered saline (TBS) containing 5% non-fat dry milk and 0.1% Tween 20 (TBST), the membranes were incubated overnight at 4°C with the respective primary antibodies, i.e. anti-P21 (Sigma-Aldrich; Cat. No.: 05–655; 1:1000 dilution) and anti-GAPDH (Merck Millipore; Cat. No.: MAB374; 1:1000 dilution). Subsequently, the membranes were washed with TBST thrice and probed with the secondary anti-mouse IgG antibody (Sigma-Aldrich; Cat. No.: NA931V; 1:5000 dilution) at room temperature for 2 h. The Clarity™ Western ECL Substrate (Bio-Rad; Cat. No.: 1705060) and the ChemiDoc Imaging System (Bio-Rad; Cat. No.: 17001402) were applied for signal detection.

***OCT4* gene tagging**

Human iPSCs were transfected with constructs designed for tagging *OCT4* through HR, HMEJ or ITPN following the scheme indicated in **Supplementary Table S22**. At 2 days post-transfection, the iPSCs were transferred to wells of 6-well plates (Greiner Bio-One) coated with VTN-N and, upon reaching approximately 50% confluency, were exposed to E8 Medium containing 0.5 μg ml^{-1} puromycin. The resulting puromycin-resistant iPSC colonies were identified by using the Leukocyte Alkaline Phosphatase Kit following the manufacturer's instructions (Sigma-Aldrich; Cat. No.: 86R-1KT). Additionally, the puromycin-resistant iPSCs were further expanded for quantifying the frequency of cells expressing OCT4::EGFP after transduction with a lentiviral vector coding for the bacteriophage P1 Cre recombinase (LV.Cre) (28,44) at a multiplicity-of-infection (MOI) of 20 vector particles per cell. The quantification of OCT4::EGFP-positive iPSCs was carried out with the aid of a BD LSR II flow cytometer (BD Biosciences).

iPSC differentiation

The *in vitro* spontaneous differentiation of iPSCs into mesoderm cells was described elsewhere (43). In brief, OCT4::EGFP$^+$ iPSCs were dissociated into large cell clumps and incubated in suspension on low-attachment plates for a period of 24 h. Afterwards, the cell clumps were replated on glass coverslips coated with Vitronectin. After two days in culture, the regular growth medium was replaced by differentiation medium, i.e., DMEM/F12 (Gibco; Cat. No. 31331-028) containing 20% FBS. The differentiation medium was replenished every 2-3 days during the following 3 weeks. The differentiation of OCT4::EGFP$^+$ iPSCs into ectoderm and endoderm cells was carried out with the aid of the STEMdiff™ Trilineage Differentiation Kit (STEMCELL Technologies; Cat. No. 05230) following the manufacturer's recommendations. Confocal immunofluorescence microscopy analyses were carried out for detecting the indicated lineage markers specific for mesoderm, ectoderm and endoderm germ layers (**Supplementary Table S36**).

Confocal immunofluorescence microscopy

OCT4::EGFP$^+$ iPSC populations were fixed with 4% paraformaldehyde (PFA), permeabilized with 0.5% Triton X-100 in tris-buffered saline (TBS) pH 7.6 (50 mM Tris-HCl pH 7.6; 150 mM NaCl), and blocked with a blocking solution consisting of TBS, Triton X-100, 2% BSA and 0.1% sodium azide. Afterwards, the cells were incubated with the corresponding primary antibodies and after thorough washes in TBS were exposed to the fluorochrome-conjugated secondary antibodies indicated in **Supplementary Table S36**. Finally, the specimens were mounted in the ProLong Gold Antifade Mounting reagent (Thermo Fisher Scientific; Cat. No.: P36931). Finally, the images were captured by using an upright Leica SP8 confocal microscope equipped with Leica hybrid detectors, HyD (Leica Microsystems) and analyzed the with the aid of the LAS X software.

Statistical analyses

Statistical analyses on data obtained from at least three biological replicates were performed with the GraphPad Prism software (version 9.3.1). Information on statistical parameters and tests used are specified in the figure legends.

RESULTS

Functional screens identify Cas9 variants with improved performance over regular Cas9 for HR and HMEJ genome editing

Gene targeting (knock-in) into safe harbor loci in single or multiplexing formats leverages and broadens synthetic biology and genetic therapy efforts (50,51). Hence, to test the performance of the different gene knock-in tools and strategies, the commonly used prototypic safe harbor loci *AAVS1* and *CCR5* were selected, together with the more recently characterized *CLYBL* locus, as endogenous target sequences (52,53). We started by comparing the performance of wild-type SpCas9:gRNA complexes with those of a representative panel of high-specificity SpCas9 variants for DSB-dependent genome editing using regular and target site-modified plasmid donors designed for HR and HMEJ, respectively. This panel consists of SpCas9 variants SpCas9-KA (8), SpCas9-KARA (8), eSpCas9(1.1) (8), Sniper-Cas9 (11), xCas9-3.7 (10), evoCas9 (9) and SpCas9-HF1 (7) (**Figure 1A**). Thus, cervical carcinoma HeLa cells were transfected with regular HR or modified HMEJ donors each mixed with isogenic constructs expressing individual nucleases and canonical gRNAs specific for *CCR5* or *AAVS1* acceptor genomic sequences (**Figure 1B**, top and bottom graphs, respectively). Of notice, in contrast to gRNAs with extended spacers and/or non-hybridizing 5' guanines, canonical *S. pyogenes* gRNAs with 20-nucleotide spacers fully complementary to protospacer DNA do not hinder high-specificity SpCas9 nuclease activities (9,11,44,54,55). After a 2-week sub-culturing period, to eliminate expression from episomal donor templates, DSB-dependent genome editing frequencies were determined by reporter-directed flow cytometry (**Figure 1B**). These experiments revealed that Sniper-Cas9 together with the nuclease sub-set formed by the single K848A, double K848A/R1060A and triple K848A/K1003A/R1060A mutants SpCas9-KA, SpCas9-KARA and eSpCas9(1.1), respectively, yielded DSB-dependent genome editing levels as high as or higher than those achieved by the parental SpCas9 nuclease (**Figure 1B** and **Supplementary Figure S2**). Indeed, frequencies reached with HR and HMEJ templates at *AAVS1* were, respectively, 13.22 ± 3.92% and 26.17 ± 3.66% when delivering SpCas9 versus 30.18 ± 6.78% and 66.14 ± 12.8% when introducing eSpCas9(1.1) (**Supplementary Figure S3**). Moreover, similarly to experiments using SpCas9, modified HMEJ-prone donors outperformed donors strictly susceptible to HR when combined with Sniper-Cas9, SpCas9-KA, SpCas9-KARA and eSpCas9(1.1) (**Figure 1B**). In contrast, genome editing frequencies induced by xCas9-3.7 and evoCas9 were lower than those triggered by SpCas9, with differences between HR and HMEJ donors not reaching significance (**Figure 1B**). Further experiments revealed that eSpCas9(1.1) outperformed SpCas9-HF1 at *AAVS1* and *CLYBL*, with the highest differences in genome editing levels reached by these two nucleases observed at the latter locus (**Supplementary Figure S4**). Specifically, DSB-dependent genome editing frequencies achieved with HR and HMEJ templates at *CLYBL* were, respectively, 0.86 ± 0.18% and 7.36 ± 2.44% when using SpCas9-HF1 versus 8.82 ± 1.52% and 54.15 ± 4.71% when deploying eSpCas9(1.1) instead (**Supplementary Figure S4**).

Subsequently, independent assays based on tracing polymorphism knock-ins in HeLa cells by restriction fragment length analysis (RFLA) (**Figure 1C**); and screening transgene knock-ins in randomly isolated iPSC colonies (n = 47) by junction PCR assays established HDR-mediated gene targeting in cells exposed to high-specificity nucleases and matched donor constructs (**Figure 1D**). Moreover, in

agreement with the experiments using AAVS1-targeting reporter constructs (**Figure 1B**, bottom graph), the RFLA assay detected the highest DSB-dependent genome editing levels when delivering the high-specificity nucleases SpCas9-KA, SpCas9-KARA and eSpCas9(1.1) together with HMEJ donor templates (**Figure 1C**).

Towards expanding the scope of HR- and HMEJ-based genome editing, we next tested the SaCas9 nuclease (**Figure 1A**) together with *AAVS1*-targeting donors in HeLa cells or with *CLYBL*-targeting donors in HeLa cells and iPSCs (**Figure 2**). DSB-dependent genome editing frequencies were measured by flow cytometry of EGFP-expressing HeLa cells or by colony-formation assays based on puromycin selection and alkaline phosphatase staining of $Puro^R$.EGFP-expressing iPSCs (**Figure 2**). In line with the experiments using SpCas9 (**Figure 1B** and **Supplementary Figure S4**), donor constructs prone to HMEJ yielded higher DSB-dependent genome editing frequencies than donors strictly susceptible to HR, independently of the cell type or genomic target region probed (Figure 2). In HeLa cells, this difference was most noticeable at *CLYBL* with HR- and HMEJ-prone donors resulting in SaCas9-edited cell frequencies of 4.77 ± 1.16% and 58.8 ± 12.19%, respectively (**Figure 2**).

Together, these experiments have identified Cas9 nucleases whose high specificities and activities turn them into preferable tools for DSB-dependent genome editing approaches. In addition, these data validate a versatile set of CRISPR reagents and matched HR- and HMEJ-tailored donor constructs for safe harbour targeting in human cells.

Functional screens identify high-specificity Cas9 variants compatible with in trans paired nicking

By enhancing otherwise inefficient SSB-dependent HR, ITPN constitutes a valuable approach for seamless chromosomal installation of large DNA segments in eukaryotic cells (24). Moreover, owing to its scarless character, ITPN is particularly useful for achieving allele-specific editing (39–41), minimizing haploinsufficiency, or for editing repetitive or essential genomic tracts (28). In addition, ITPN has been applied for one-step biallelic and multiplexing DNA editing and for clonal screening-free generation of model cells and organoids (24,41).

Previous research from our laboratory using DNA/gRNA mismatch screens demonstrated that the specificities of mutant $SpCas9^{D10A}$ variants exceeds by manifold that of the parental $SpCas9^{D10A}$ nickase (44). Here, to further improve the seamless and scarless character of ITPN genome editing, we sought to investigate its compatibility with these high-specificity nickases, namely, $SpCas9\text{-}KA^{D10A}$, $SpCas9\text{-}KARA^{D10A}$, $eSpCas9(1.1)^{D10A}$, $Sniper\text{-}Cas9^{D10A}$, $xCas9\text{-}3.7^{D10A}$, $evoCas9^{D10A}$ and $SpCas9\text{-}HF1^{D10A}$ (**Supplementary Figure S5**). To this end, we started by comparing the performances of parental SpCas9D10A:gRNA complexes with those of high-specificity $SpCas9^{D10A}$ variants using regular and target site-modified donors for single nicking (SN)- and ITPN-mediated HR, respectively. Thus, HeLa cells were transfected with unmodified or target site-modified donors together with isogenic constructs expressing specific nickases and canonical gRNAs targeting *CCR5* and *AAVS1* acceptor sequences (**Figure 3A**, top and bottom graphs, respectively). After a 2-week sub-culturing period, SSB-dependent genome editing frequencies were assessed by reporter-directed flow cytometry. These experiments revealed that, at *CCR5*, the best-performing nickase was $SpCas9\text{-}KA^{D10A}$ (**Figure 3A**, top graph) whilst at *AAVS1*, $SpCas9\text{-}KA^{D10A}$ together with $SpCas9\text{-}KARA^{D10A}$ and $eSpCas9(1.1)^{D10A}$ induced ITPN genome editing to the same extent as the parental $SpCas9^{D10A}$ nickase (**Figure 3A**, bottom graph). Consistent with the nuclease screens (**Figure 1**), $xCas9\text{-}3.7^{D10A}$ and $evoCas9^{D10A}$ triggered the lowest frequencies of SSB-dependent genome editing (**Figure 3A**). Additional experiments showed that both $SpCas9^{D10A}$ and $eSpCas9(1.1)^{D10A}$ outperformed $SpCas9\text{-}HF1^{D10A}$ at *AAVS1* and *CLYBL*, with the highest ITPN genome editing levels induced by these nickases registered at the former locus (**Supplementary Figure S6**).

Significantly, the comparison of precise HR setups encompassing ITPN and genomic DSBs (canonical HR), revealed that, except when directing $eSpCas(1.1)^{D10A}$ to *AAVS1*, ITPN reached similar or significantly higher frequencies of genome-edited cells than canonical HR (**Figure 3B**). Complementing *AAVS1* gene targeting experiments in iPSCs using SpCas9 and $eSpCas9(1.1)^{D10A}$, besides confirming the poor performance of SN genome editing (**Figure 3C**), further corroborated that ITPN mostly avoids target allele disruptions (**Figure 3D**) while achieving precise HR-derived chromosomal insertions (**Figure 3E**).

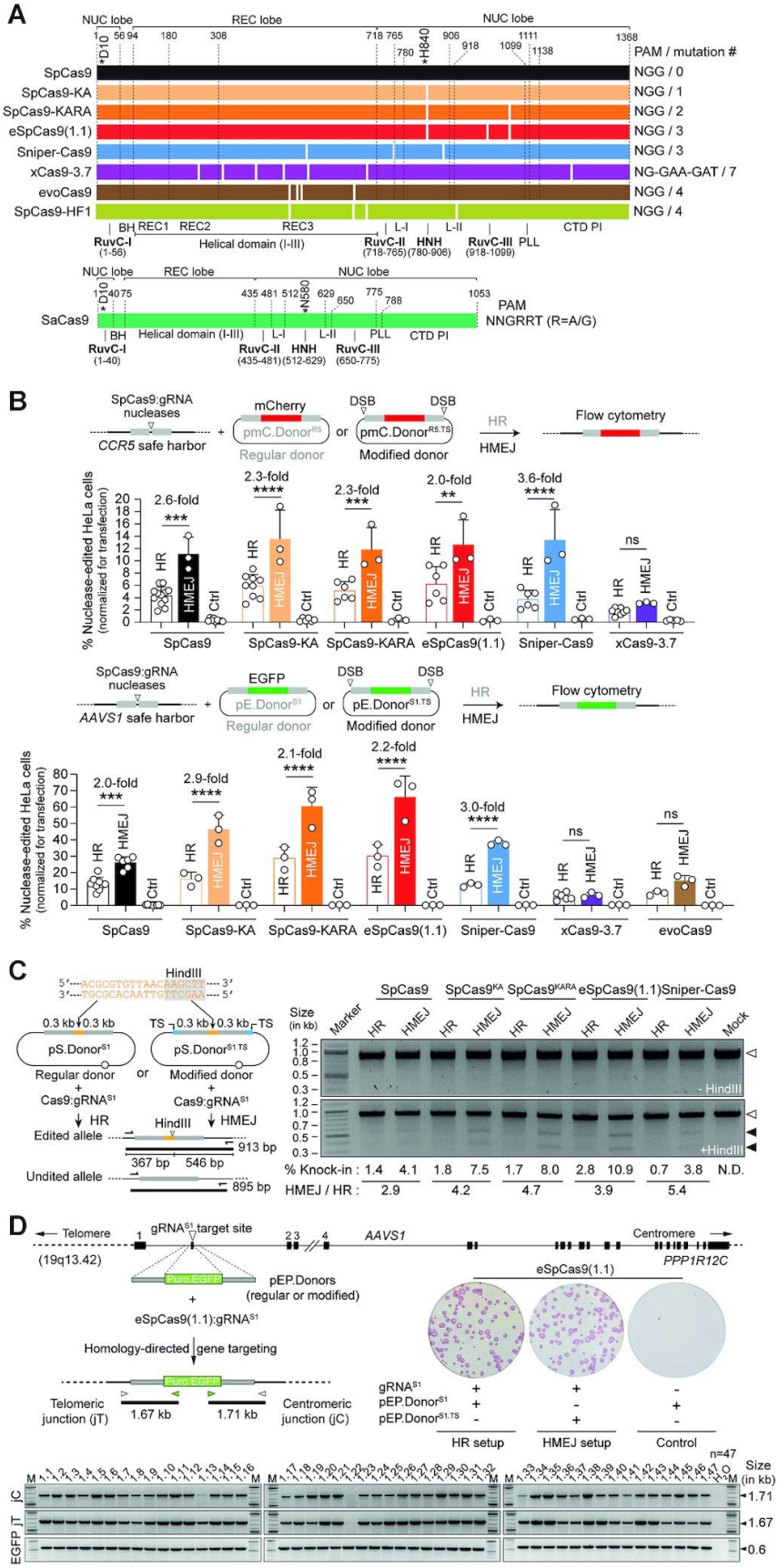


Figure 1. Testing DSB-dependent genome editing using regular versus high-specificity SpCas9 nucleases. (**A**) Diagrams of engineered Cas9 nucleases derived from *S. pyogenes* and *S. aureus* type II CRISPR systems. Protein domains and mutation positions are marked by dashed and white lines, respectively. HNH, histidine-asparagine-histidine nuclease domain; RuvC, RuvC-like nuclease domain composed of a tripartite assembly of RuvC-I, -II and -III. The HNH and RuvC domains in the nuclease lobe cut the target and non-target DNA strands, respectively. L-I and L-II, linker region I and II, respectively. BH, Arginine-rich bridge

helix; CTD, C-terminal domain in which the PAM-interacting motif (PI) is integrated; NUC and REC, nuclease and recognition lobes, respectively; PLL, phosphate lock loop. Asterisks mark residues D10 and H840 crucial for RuvC and HNH catalytic activities, respectively. (**B**) Genome editing based on donors prone to canonical HR and HMEJ upon high-specificity SpCas9 delivery. Nuclease-dependent genome editing frequencies in HeLa cells transfected with the depicted reagents targeting *AAVS1* and *CCR5* were quantified by reporter-directed flow cytometry at 17 days post-transfection (top and bottom graphs, respectively). HeLa cells exposed to corresponding Cas9 nucleases and regular donor plasmids in the absence of locus-specific gRNAs served as negative controls. Data are plotted as mean ± SD of at least 3 independent biological replicates. Significant differences between the indicated datasets were determined by two-way ANOVA followed by Šidák's multiple comparisons tests; ****P<0.0001, ***0.0001<P<0.001, **0.001<P<0.01; P> 0.05 was considered non-significant (ns). (**C**) Genotyping assay assessing HDR-mediated restriction site knock-ins. Regular pS.Donor^{S1} and modified pS.Donor$^{S1.TS}$ constructs, designed to introduce a HindIII recognition site at *AAVS1* through HR and HMEJ processes, respectively, were transfected into HeLa cells together with plasmids expressing SpCas9 nucleases and gRNA^{S1}. The HindIII polymorphism is detected by restriction-fragment length analysis (RFLA) of amplicons covering the target site (left panel). RFLA products diagnostic for unedited and edited *AAVS1* alleles retrieved from HeLa cells exposed to the indicated reagents were measured through densitometry and are marked with open and closed arrowheads, respectively (right panel). (**D**) Genotyping assay assessing HDR-mediated transgene knock-ins. Regular pEP.Donor^{S1} and modified pEP.Donor$^{S1.TS}$ plasmids, tailored for inserting the live-cell selectable marker EGFP::Puro^{R} at *AAVS1* via HR and HMEJ processes, respectively, were transfected into iPSCs together with constructs expressing eSpCas9(1.1):gRNA^{S1} complexes. HDR-derived gene knock-ins were identified by junction PCR analysis of randomly selected iPSC clones engineered through pEP.Donor^{S1} and eSpCas9(1.1):gRNA^{S1} delivery. Puromycin-resistant iPSC colonies were identified by staining for the pluripotency marker alkaline phosphatase.

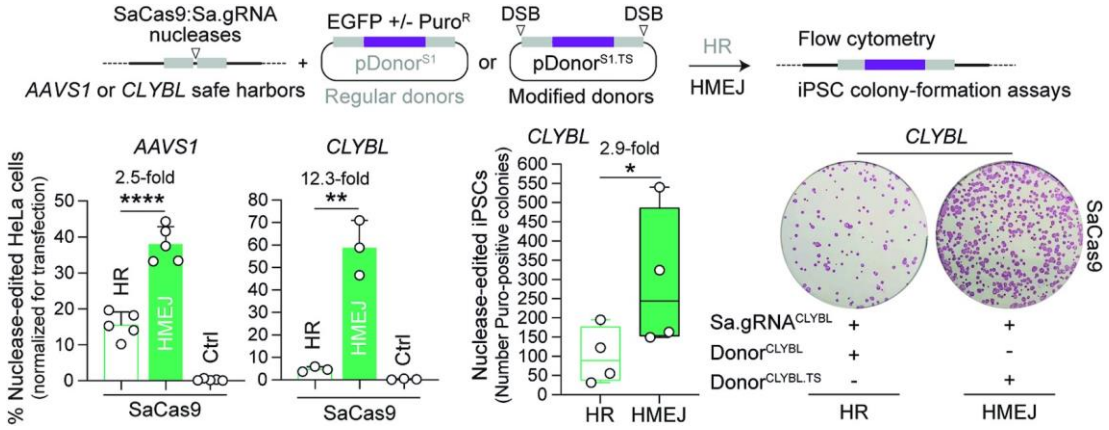


Figure 2. Genome editing combining plasmid donors with regular HR or modified HMEJ templates and orthogonal SaCas9 complexes. SaCas9-dependent genome editing at *AAVS1* and *CLYBL* loci in HeLa cells using EGFP-encoding donors, and at *CLYBL* in iPSCs using Puro^{R}.EGFP-encoding donors was determined by reporter-directed flow cytometry and colony-formation assays, respectively. The latter assays detected the pluripotency marker alkaline phosphatase to identify puromycin-resistant iPSCs. Controls consisted of cells exposed to regular donor plasmids and SaCas9 nucleases with non-targeting gRNAs. Data are presented as mean ± SD of at least three independent biological replicates. Significant differences between the indicated datasets were calculated by two-tailed unpaired Student's *t* tests (left and middle graphs) and two-tailed paired ratio t test (right graph); ****P < 0.0001, **0.001< P< 0.01, *0.01< P< 0.05.

Towards broadening the scope of ITPN genome editing, we next performed experiments in HeLa cells and iPSCs using *AAVS1*- and *CLYBL*-targeting donors together with orthologue SaCas9^{D10A} and SaCas9^{N580A} nickases (**Figure 4**). SSB-dependent genome editing frequencies were assessed by flow cytometry of EGFP-expressing HeLa cells or by iPSC colony-formation assays (**Figure 4**). Consistent with the experiments using parental SpCas9^{D10A} and high-specificity SpCas9^{D10A} derivatives (**Figure 3A**), the HR setups involving ITPN were more effective than those entailing SN (**Figure 4**). However, in contrast to the experiments using SpCas9^{D10A} nickases (**Figure 3A**), neither SaCas9^{D10A} nor SaCas9^{N580A} led to genome editing frequencies higher than those obtained through SaCas9-induced canonical HR (**Figure 4** and **Supplementary Figure S7**). These data indicate that when compared to *S. aureus* SaCas9^{D10A} and SaCas9^{N580A} nickases, *S. pyogenes* SpCas9^{D10A} nickases are preferable for ITPN genome editing, especially so in their high-specificity configurations.

Orthogonal high-throughput genome-wide translocation sequencing (oHTGTS) permits tracing off-target effects of CRISPR nucleases vis-à-vis nickases in a quantitative and unbiased fashion (28,44). In our earlier work, oHTGTS assays showed a striking and progressive reduction of off-target activities associated with SpCas9, high-specificity eSpCas9(1.1) and SpCas9^{D10A}. A more moderate, yet readily measurable, further reduction in off-target effects was detected when using the high-specificity eSpCas9(1.1)^{D10A} nickase instead of its parental SpCas9^{D10A} counterpart (44). Moreover, oHTGTS assays also disclosed sequences mapping at *CPNE5* and *BBOX1* as the top-ranked off-target sites for CRISPR complexes formed by coupling the *AAVS1*-specific gRNA^{S1} to SpCas9 and SpCas9^{D10A},

respectively (28). Therefore, we proceeded by assessing the integrity of *AAVS1*, *CPNE5* and *BBOX1* in HeLa cell populations genome-edited through canonical HR using SpCas9:gRNA^{S1}, eSpCas9(1.1):gRNA^{S1} or Sniper-Cas9:gRNA^{S1} or via ITPN using their corresponding D10A nickase derivatives (**Figure 5A**). Targeted amplicon deep sequencing confirmed high and low indel frequencies at *AAVS1* in cells exposed to nuclease and nickase complexes, respectively (**Figure 5B**). Furthermore, in striking contrast to eSpCas9(1.1), the regular SpCas9 and high-specificity Sniper-Cas9 nucleases led to similar and high frequencies of indels at the *CPNE5* off-target site. Significantly, none of the nickase complexes tested induced detectable off-target activities using the sensitive deep sequencing genotyping assays (**Figure 5B**).

As aforementioned, 'double-cut' donors susceptible to HMEJ, MMEJ and NHEJ are normally more efficient genome editing substrates than their HR counterparts. Yet, the free termini generated in cellula from 'double-cut' donors upon site-specific DNA cleavage might diminish the genome editing precision due to end-to-end ligation ('capture') at off-target DSBs (24,56). Thus, to further investigate genome editing precision using conventional and high-specificity Cas9 proteins, HeLa cells were genetically modified through HR, HMEJ and ITPN (**Figure 6A**), and then analysed for on-target and off-target donor DNA insertion at *AAVS1* and *CPNE5* (**Figure 6B** and **C**, respectively). Besides confirming donor DNA targeting at *AAVS1* (**Figure 6B**), junction PCR analysis established that HMEJ donors are the most prone to HR-independent 'capture' at off-target sequences and that these unwanted outcomes can be minimized by using high-specificity instead of parental SpCas9 nucleases (**Figure 6C**). Finally, genotyping assays based on T7 endonuclease I (T7EI) digestions for indel detection (**Figure 6C**) and DNA sequencing (**Supplementary Figure S8**) strengthened the value of ITPN for precise chromosomal insertion of large genetic payloads with minimal bystander effects at target alleles within genome-edited cell populations.

Collectively, these experiments have identified Cas9 nickases whose combined high specificities and activities turns them into valuable alternatives to the regular SpCas9^{D10A} for use in ITPN genome editing settings and stress the relevance of using high-specificity SpCas9 nucleases, especially when aiming at targeted insertion of free-ended donor DNA.

CRISPR-Cas9 nickases fail to activate the P53-dependent DNA damage response in iPSCs

Single to few DSBs suffice to induce P53 signalling in stem cells (29,30) causing cell cycle arrest at G1. Hence, CRISPR-Cas9-induced HR is hindered in cells with functional P53 as it takes place during the S thorough G2 phases of the cell cycle (25). Indeed, P53 absence or inhibition correlates with enhanced DSB-dependent genome editing (29-31,57).

A recent study showed that SpCas9^{D10A} did not significantly activate P53 signalling in cervical carcinoma and mammary epithelial cell lines, i.e. HeLa and MCF10A cells, respectively (37). To examine P53 signalling elicited by nickases versus nucleases in cells with a low sensitivity threshold to DNA damage, we selected human iPSCs owing to their established relevance in basic and translational research. Besides present in over 50% of cancers, cells with P53 mutations can recurrently arise in cultures of pluripotent stem cells (PSCs) (58). Importantly, real-time cell proliferation assays in the presence and absence of Nutlin-3a, a small-molecule inhibitor of P53-MDM2 interactions (**Figure 7A**), demonstrated that the selected iPSCs have a functional P53 status (**Figure 7B** and **C**; **Supplementary files S1** and **S2**). This conclusion was independently confirmed by measuring cell viability using metabolic and apoptosis activity assays (**Figure 7D** and **E**, respectively) and by detecting the specific upregulation of the P53 target genes *FAS*, *PUMA* and *MDM2* at the transcriptional level and of P21 at the transcriptional and protein levels (**Figure 7F** and **G**, respectively).

Next, the iPSCs were transfected with constructs expressing regular or high-specificity SpCas9 proteins and gRNA^{CALM2} or gRNA^{VEGFA}. The former and latter gRNAs are known to have few and numerous off-target sites, respectively, as assessed in silico (**Supplementary Figure S9**) and experimentally (8,29,45). Expression analysis of the P53 transcription factor target genes *FAS* and *P21* disclosed that coupling SpCas9 and high-specificity eSpCas9(1.1) nucleases to the promiscuous gRNA^{VEGFA} led to significant activation of P53 signaling, whilst coupling the same gRNA^{VEGFA} to SpCas9^{D10A} and eSpCas9(1.1)^{D10A} nickases, did not (**Figure 8A**). Moreover, high-specificity gRNA^{CALM2} also led to nuclease-dependent upregulation of *FAS* and *P21* expression (**Figure 8A**). Cumulative datasets

comparing nuclease- versus nickase-mediated activation of the P53-responsive genes *FAS*, *P21*, *PUMA* and *MDM2* revelated that SpCas9^{D10A} nickases are poor triggers of the P53-dependent DDR in iPSCs when compared to SpCas9 nucleases (**Figure 8B**).

Together, these results indicate that genome editing with SpCas9^{D10A} nickases might offer a heightened safety profile to engineered cell products derived from iPSCs in that, besides cell-cycle arrest and apoptosis, DSB-induced signalling pathways have been linked to the selection of cells with mutations in cancer-associated genes, i.e. TP53 and KRAS (31,32).

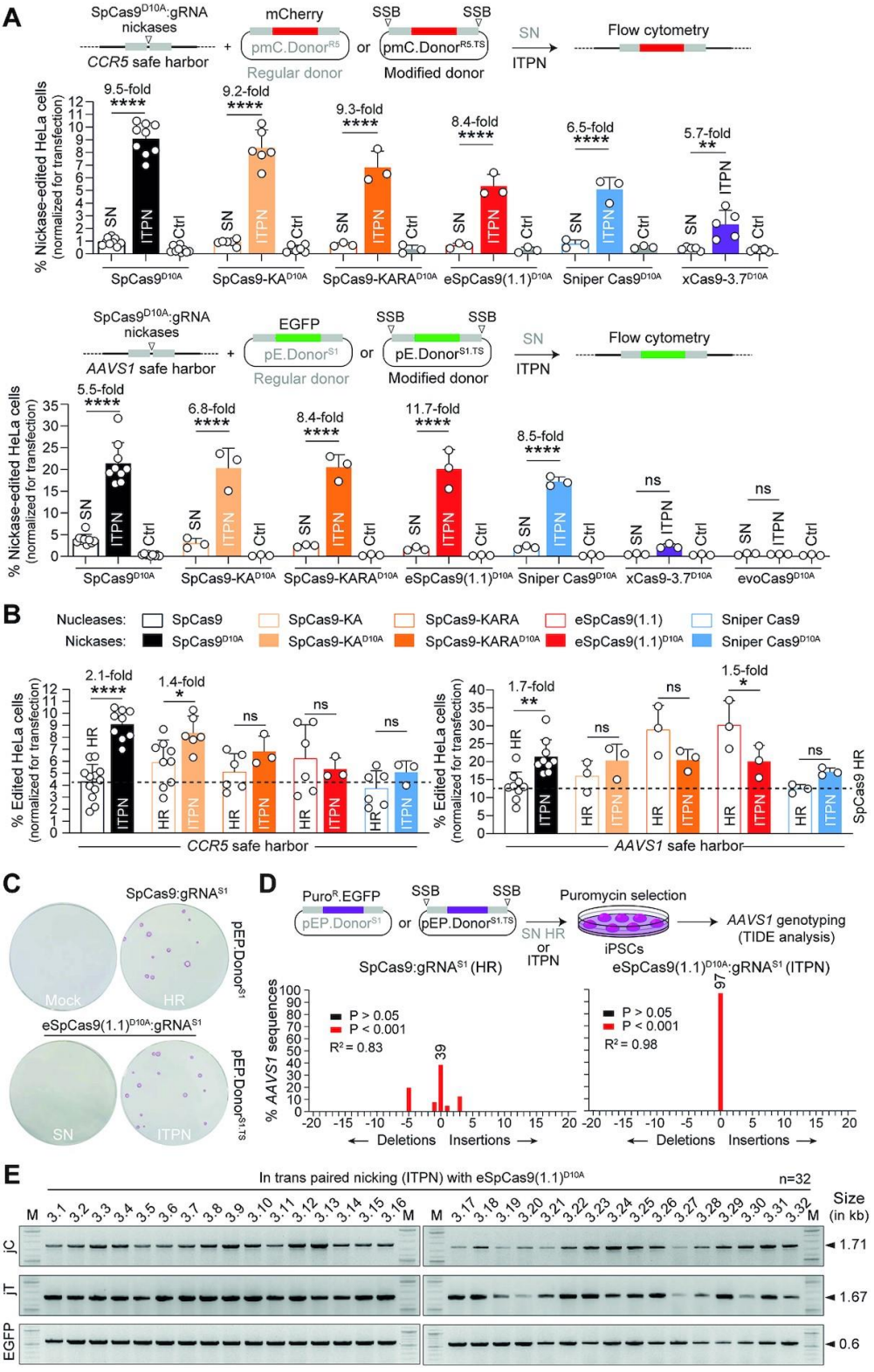


Figure 3. Testing SSB-dependent genome editing using regular versus high-specificity SpCas9^{D10A} nickases. (**A**) Single nicking and in trans paired nicking genome editing based on high-specificity SpCas9^{D10A} variants. Nickase-dependent genome editing frequencies in HeLa cells transfected with the depicted components targeting *CCR5* and *AAVS1* were measured by reporter-directed flow cytometry at 17 days post-transfection (top and bottom graphs, respectively). HeLa cells treated with corresponding Cas9^{D10A} nickases and regular donor plasmids in the absence of locus-specific gRNAs served as negative controls. Results are plotted as mean ± SD of at least three independent biological replicates. Significant differences between the indicated datasets were assessed by two-way ANOVA followed by Šidák's multiple comparisons test; ****P< 0.0001, **0.001<P<0.01; P>0.05 was considered non-significant (ns). (**B**) Comparing standard and in trans paired nicking genome editing strategies at *CCR5* and *AAVS1*. Plotting of datasets presented in panel A corresponding to HeLa cells subjected to nucleases and regular donors or to nickases and target site-modified donors (canonical HR or ITPN strategies, respectively). Dashed lines mark the means of the DSB-dependent genome editing levels obtained with conventional SpCas9 and unmodified HR donor templates. Data are shown as mean ± SD of at least 3 independent biological replicates. Significant differences between the indicated datasets were calculated by two-way ANOVA followed by Šidák's multiple comparisons tests; ****P<0.0001, **0.001<P<0.01, *0.01<P<0.05; P>0.05 was considered non-significant (ns). (**C**) Testing standard and in trans paired nicking in iPSCs using high-specificity cleaving and nicking CRISPR complexes. iPSCs edited upon exposure to the indicated *AAVS1*-targeting reagents were selected in the presence of puromycin and the resulting colonies were stained for the pluripotency marker alkaline phosphatase. (**D**) Probing mutagenic loads in genome-edited iPSCs. iPSCs edited after exposure to the indicated *AAVS1*-targeting reagents were selected in the presence of puromycin and indel profiles at *AAVS1* were examined through tracking of indels by decomposition (TIDE) analysis. (**E**) Establishing HDR-mediated transgene insertion in iPSCs edited through in trans paired nicking. Junction PCR analysis was performed on randomly picked iPSC clones engineered through pEP.Donor$^{S1.TS}$ and eSpCas9(1.1)^{D10A}:gRNA^{S1} delivery.

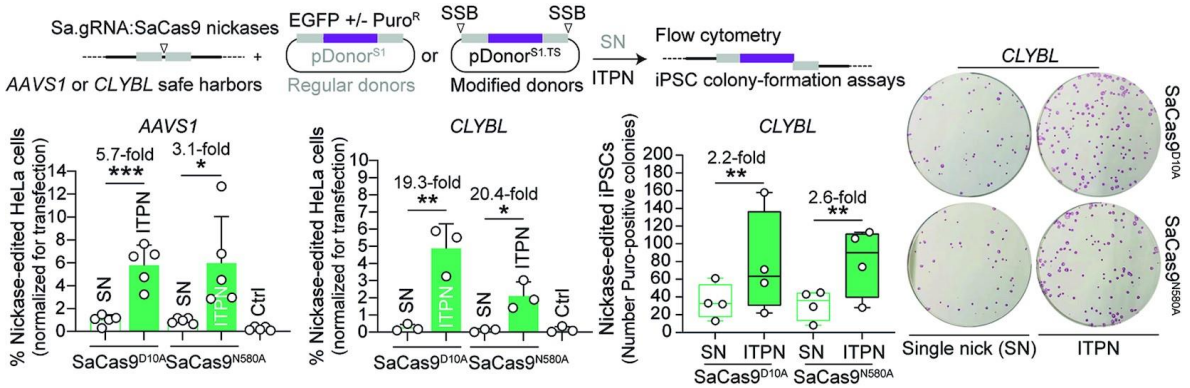


Figure 4. Genome editing combining regular SN plasmid donors or modified ITPN donors and nicking orthogonal SaCas9 complexes. SaCas9^{D10A}- or SaCas9^{N580A}-dependent genome editing at *AAVS1* and *CLYBL* loci in HeLa cells using EGFP-encoding donors, and at these loci in iPSCs using Puro^R.EGFP-encoding donors, was assessed by reporter-directed flow cytometry and colony-formation assays, respectively. The latter assay detected the pluripotency marker alkaline phosphatase to identify puromycin-resistant iPSCs. Controls consisted of cells exposed to regular donor plasmids and nickases lacking locus-specific gRNAs. Data are shown as mean ± SD of at least three independent biological replicates. Significant differences between the indicated datasets were calculated by two-tailed unpaired Student's *t* tests (left and middle graphs) and two-tailed paired ratio *t* test (right graph); ***0.0001<P<0.001, **0.001<P<0.01, *0.01<P<0.05.

In contrast to genome editing based on high-specificity Cas9 nucleases, ITPN facilitates editing essential and non-unique allelic sequences in iPSCs

Programmable nucleases can elicit cell fitness losses and unpredictable phenotypes upon cutting DNA sequences coding for essential proteins or motifs or that are recurrent in the genome (28,59,60). *OCT4* (alias *POU5F1*) encodes a transcription factor essential for human embryogenesis (61) and PSC maintenance (62,63). The essentiality of *OCT4* combined with its extensive homology to pseudogenes *POU5F1B*, *POU5F1P3*, *POU5F1P4* and *POU5F1P5* makes its editing particularly challenging. Indeed, at both coding termini, *OCT4* shares 100% homology to pseudogene sequences making it impossible designing gRNAs specific for these regions or for tagging OCT4 (**Supplementary Figure S10**). Hence, retrieving PSCs edited at such multiple-copy sequences is expected to be hindered by the acute sensitivity of these cells to DSBs. Three lines of evidence support this assertion. Firstly, OCT4 tagging experiments in iPSCs using TALENs and donor construct pDonor^{OCT4} (**Figure 9A**), did not yield HR-targeted clones (0/48) (64). Secondly, experiments in human embryonic stem cells deploying SpCas9 and donor templates with the same 'homology arms' present in pDonor^{OCT4}, led only to eight HR-targeted clones (8/288) (65). Finally, our earlier experiments in iPSCs showed that, in contrast to pDonor^{OCT4} and SpCas9:gRNA$^{OCT4.1}$ delivery (HR setup), transfer of modified pDonor$^{OCT4.TS}$ and nicking SpCas9^{D10A}:gRNA$^{OCT4.1}$ complexes (ITPN setup), readily led to OCT4-tagged iPSC populations from which viable HR-targeted iPSC clones were obtained (21/22) (28).

In this study, complementing experiments using the same live-cell gene editing readout and high-specificity DNA cleaving and nicking CRISPR complexes revealed that although canonical HR, HMEJ and ITPN setups all led to stably transfected iPSCs (**Figure 9B**), only the latter setup resulted in accurate *OCT4* editing at frequencies significantly above background levels (**Figure 9C**). These results demonstrate that despite high-specificity Cas9 nuclease usage, generating viable OCT4-tagged iPSCs is nonetheless hindered when applying the DSB-dependent genome editing strategies. Importantly, this is not the case when employing the ITPN approach instead. Moreover, dual-colour confocal microscopy analysis established that iPSCs edited through eSpCas9(1.1)^{D10A}-induced ITPN contained engineered OCT4::EGFP fusion proteins properly localized in cell nuclei (**Figure 9D**). Finally, the *OCT4* edited cells were capable of differentiating into cells representing the three embryonic germ layers, i.e. endoderm, mesoderm and ectoderm (**Figure 9E**).

Collectively, these results support the proposition that, irrespective of their specificities, programmable nucleases are outperformed by nickases for targeted and high-fidelity DNA knock-ins at sequences associated with essentiality and recurrence in the genome.

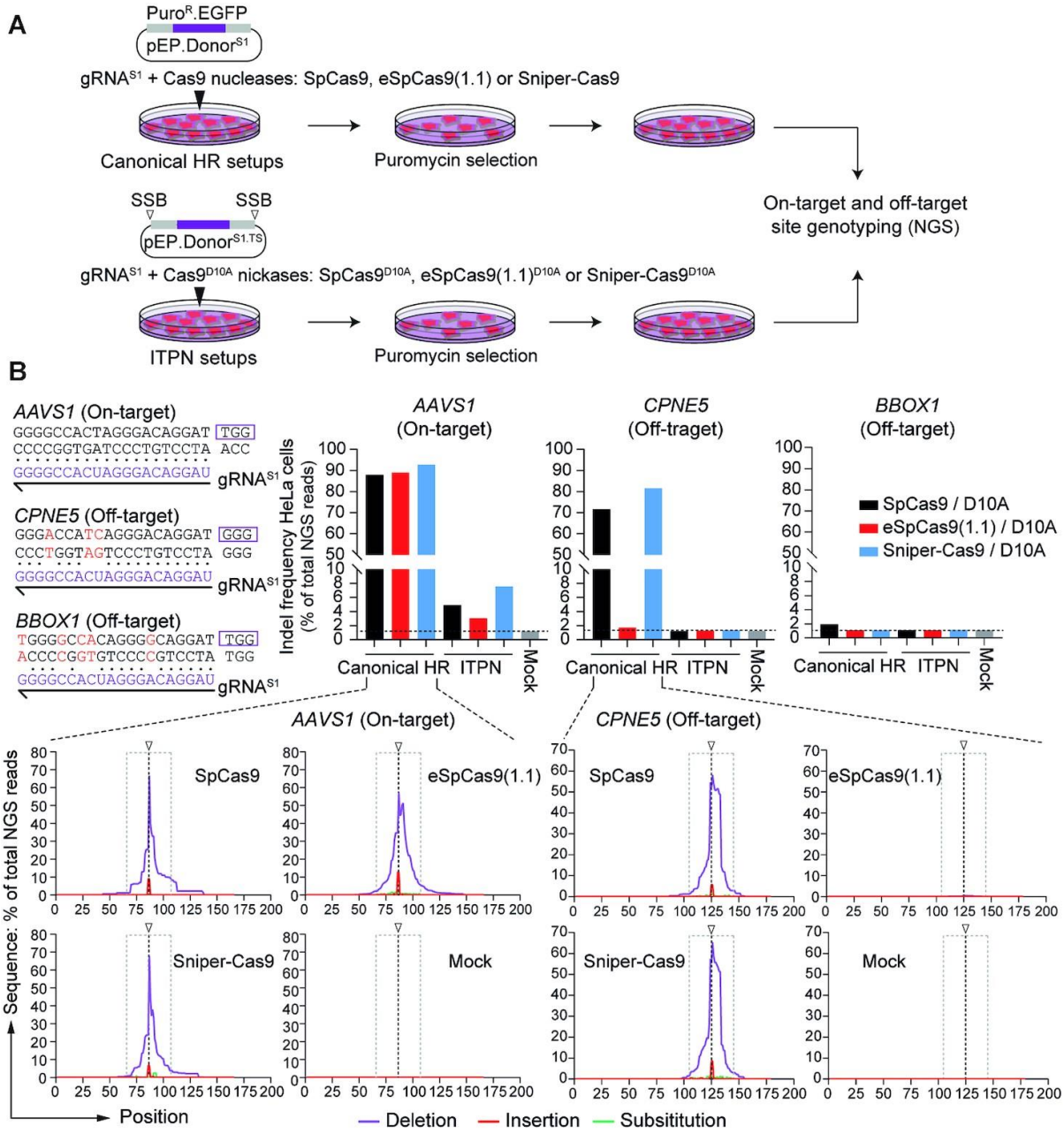


Figure 5. Assessing mutagenic loads in cells edited through canonical homologous recombination versus in trans paired nicking. (**A**) Experimental design. HeLa cells were exposed to regular and modified donors conferring puromycin resistance

together with SpCas9 nucleases and SpCas9^{D10A} nickases, respectively. SpCas9, eSpCas9(1.1) and Sniper-Cas9 nucleases, and their D10A nickase derivatives, were coupled to *AAVS1*-targeting gRNA^{S1}. Indel frequencies at on-target and off-target sites was done by amplicon deep sequencing genotyping of puromycin-resistant cell populations. (**B**) Quantification of indels at on-target and off-target sites. CRISPR complex-derived indels at the *AAVS1* target site and at two validated off-target sites (i.e. *CPNE5* and *BBOX1*) were quantified by amplicon deep sequencing (~100,000 paired-end reads per sample). Nucleotide mismatch positions between gRNA^{S1} spacer and off-target *CPNE5* and *BBOX1* sequences are highlighted in red. The types and distributions of indels detected within *AAVS1*, *CPNE5* and *BBOX1* in cells treated with regular and high-specificity nucleases are plotted. HeLa cells not exposed to CRISPR complexes provided for negative controls (Mock).

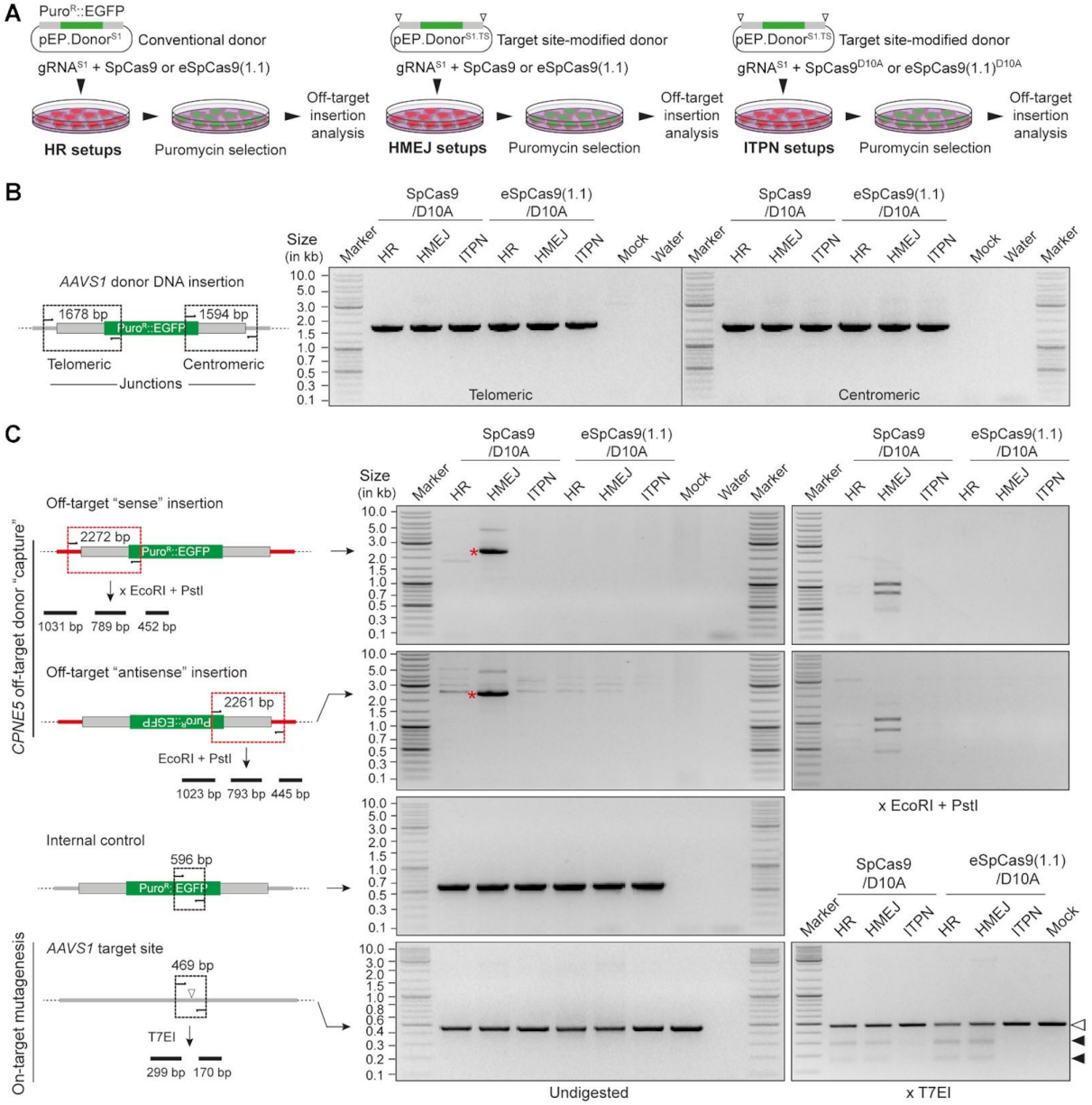


Figure 6. Assessing off-target chromosomal donor DNA insertions resulting from HR, HMEJ and ITPN using regular and high-specificity Cas9 enzymes. (**A**) Experimental design. HeLa cells were subjected to HR, HMEJ and ITPN procedures using the indicated combinations of donor DNA constructs and Cas9 proteins coupled to *AAVS1*-targeting gRNA^{S1}. Genetically modified cells, selected through puromycin exposure, were screened for donor DNA 'capture' at the prevalent gRNA^{S1} off-target site *CPNE5* by junction PCR analysis. (**B**) On-target donor DNA insertion analysis. Amplicons diagnostics for HDR-mediated *AAVS1* knock-ins are illustrated and shown. (**C**) Off-target insertion and on-target mutagenesis analysis. Amplicons diagnostics for HDR-independent 'capture' of donor DNA sequences at *CPNE5* in the 'sense' and 'antisense' orientations are illustrated and marked with asterisks. Specific donor DNA 'capture' at *CPNE5* off-target alleles and mutagenesis at *AAVS1* target alleles were probed via restriction enzyme (EcoRI and PstI) and T7 endonuclease I (T7EI) digestions, respectively. Solid arrowheads point to T7EI-digested products derived from indel-containing *AAVS1* sequences. PCR amplifications of a 596-bp *EGFP* tract served as internal controls.

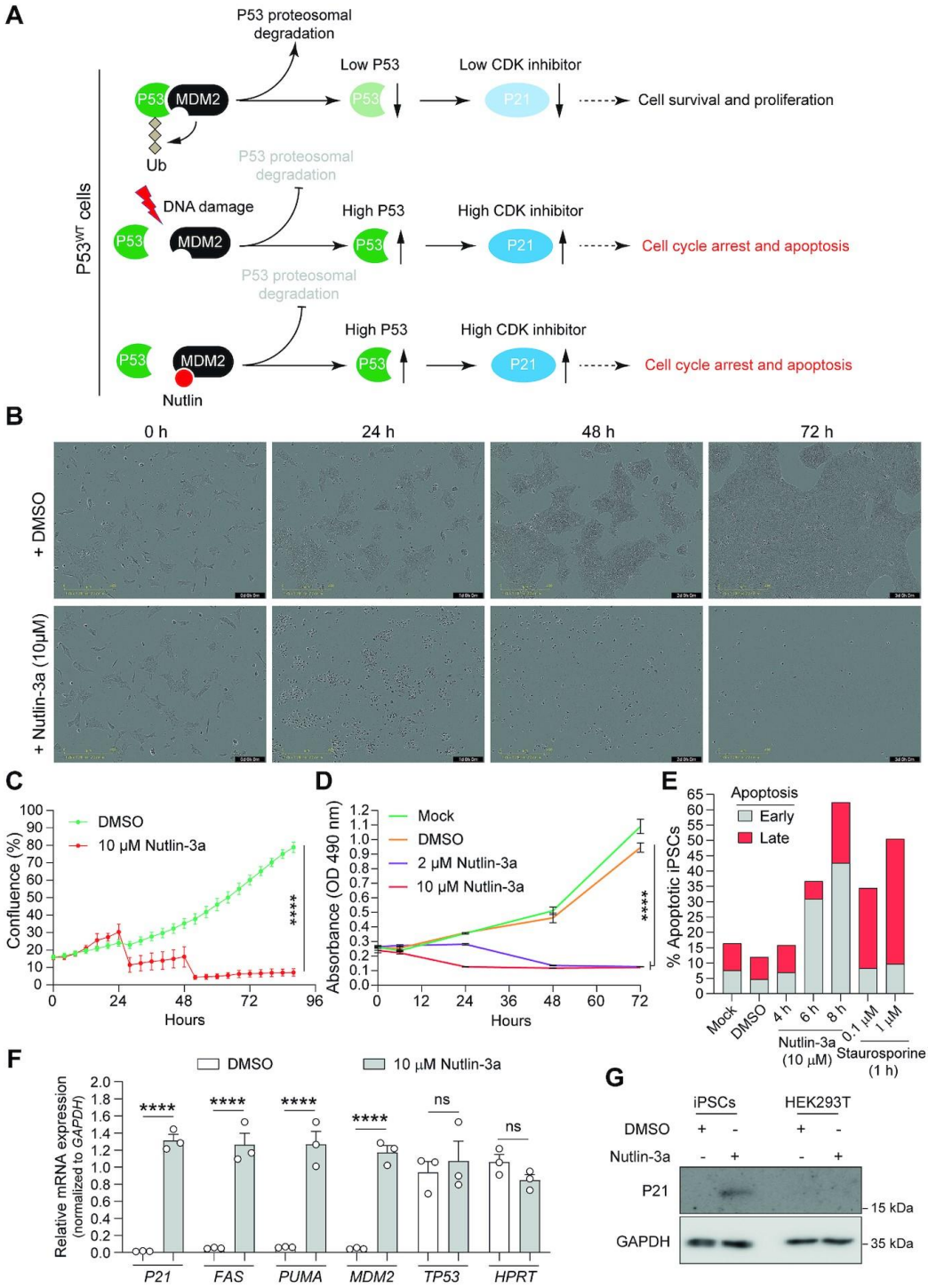


Figure 7. Cell survival assay for assessing P53 functionality in human iPSCs. (**A**) Schematics of post-transcriptional P53 activity control by DNA damage and Nutlins. In cells with normal amounts of P53, DNA damage activates ATM/ATR kinases that disrupt P53-MDM2 interaction through P53 phosphorylation. Free P53 escapes proteasomal degradation and upregulates the expression of downstream target genes (e.g. cyclin-dependent kinase inhibitor P21) inducing cell cycle arrest and apoptosis. Nutlins disrupt the P53-MDM2 interaction by instead occupying the P53 binding pocket in MDM2 mimicking a P53-dependent DNA damage response. Conversely, in cells with no or low amounts of P53, nutlins induce neither cell cycle arrest nor apoptosis (not drawn). (**B** and **C**) Realtime cell proliferation assay. The proliferation of human iPSCs incubated in the presence of Nutlin-3a (10 μM) or vehicle (DMSO) was quantified in a live-cell imaging system (IncuCyte) for 3 days. Data are shown as mean ± SD of 6 technical replicates. Significant differences between the indicated datasets were calculated by two-way ANOVA tests; ****P<0.0001. (**D** and **E**) Cell survival assays. The survival of human iPSCs incubated in regular medium (Mock) or in medium supplemented with DMSO or Nutlin-3a (2 μM and 10 μM) was monitor for 3 days by using the MTS cell metabolic activity readout (panel D). The frequencies of apoptotic human iPSCs were determined with a combined annexin V/propidium iodide assay (panel E). Annexin V positive cells and annexin V/propidium iodine doubly positive cells measured by flow cytometry scored for early and late apoptosis, respectively. Prior to flow cytometry the cells were incubated in regular medium (Mock) and in medium supplemented with DMSO or with Nutlin-3a (10 μM) for different periods. Staurosporine applied at the indicated conditions served

as an apoptosis-inducing control. (**F**) Assessing P53-dependent responses in human iPSCs exposed to Nutlin-3a. RT-qPCR analysis of transcripts for P53 and P53-responsive genes were conducted in human iPSCs incubated for 5 h in regular medium or in medium supplemented with Nutlin-3a (10 μM). RT-qPCR analysis of *HPRT1* transcripts served to measure the expression of a P53-independent control gene (n = 3 independent biological replicates). Significances were calculated with two-way ANOVA followed by Šidák's test for multiple comparisons; ****P<0.0001; P>0.05 was considered non-significant (ns). (**G**) P53-dependent P21 protein detection assay. Western blot analysis of P21 expression in human iPSCs incubated in the presence of Nutlin-3a (10 μM) or vehicle (DMSO) for 12 h. Transformed P53-defective HEK293T cells exposed to the same experimental conditions served as control. Western blotting of the housekeeping GAPDH provided for loading controls.

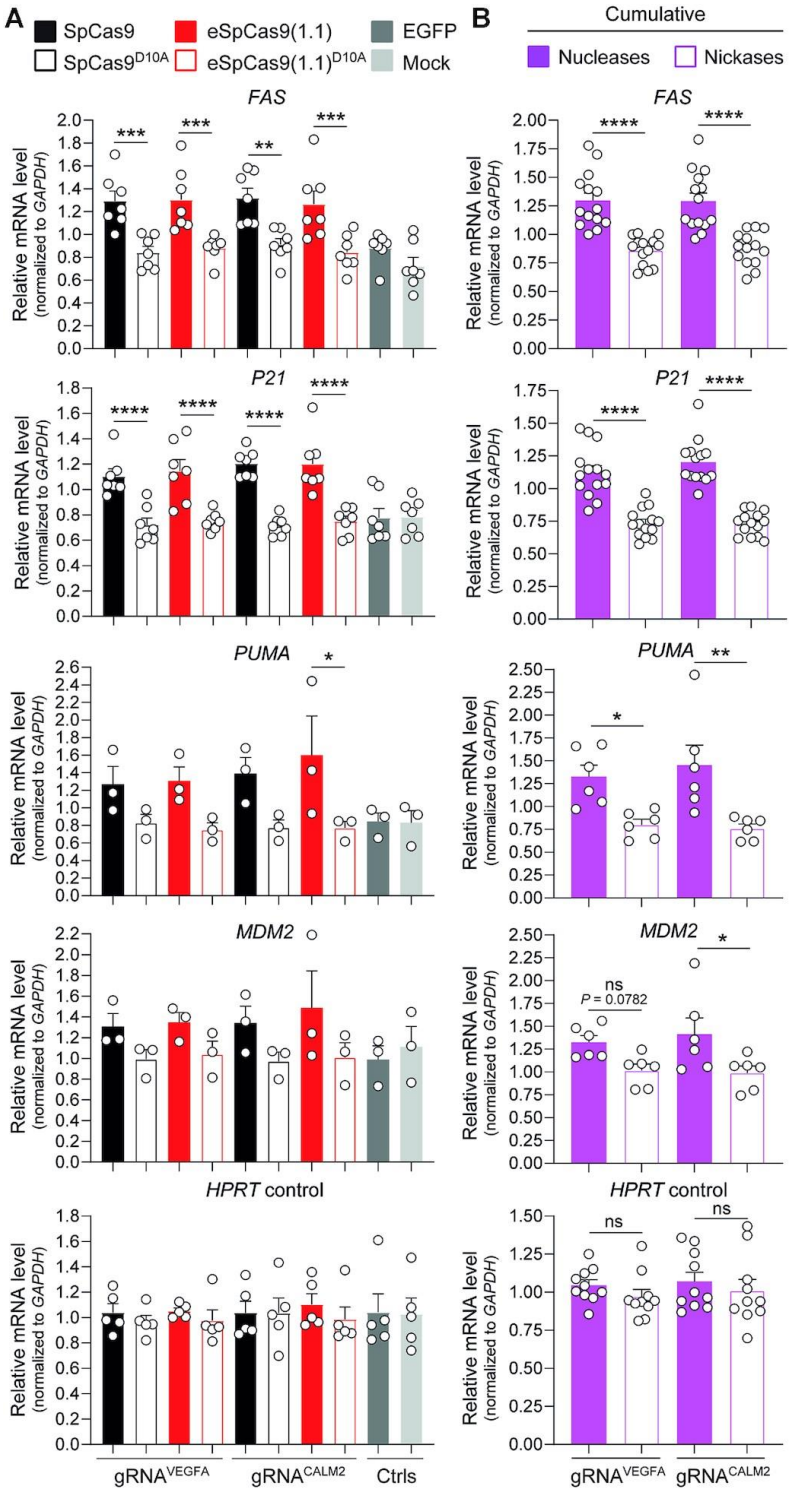


Figure 8. Assessing activation of P53-dependent DNA damage responses in human iPSCs exposed to nucleases versus nickases. (**A**) Expression analysis of P53 activation-responsive genes. Constructs encoding the indicated Cas9 enzymes and gRNAs conferring high (gRNA^{VEGFA}) or low (gRNA^{CALM2}) off-target activities (**Supplementary Figure S9**), were transfected into iPSCs. RT-qPCR measurements of *FAS*, *P21*, *PUMA* and *MDM2* transcripts whose expression is upregulated upon P53 activation

(minimum n = 3 independent biological replicates). Targeting *HPRT1* transcripts served for RT-qPCR measurements of a housekeeping control gene (n = 5 independent biological replicates). Additional controls consisted of targeting *FAS*, *P21*, *PUMA*, *MDM2* and *HPRT1* transcripts in mock-transfected iPSCs and in iPSCs transfected with an EGFP-encoding plasmid. Significances were calculated with one-way ANOVA followed by Tukey's test for multiple comparisons; ****P<0.0001, ***0.0001<P<0.001, **0.001<P< 0.01, *0.01<P<0.05. (**B**) Cumulative comparison of cleaving versus nicking effects on P53-responsive gene modulation. Combined RT-qPCR datasets derived from iPSCs treated with nucleases SpCas9 and eSpCas9(1.1) or nickases SpCas9^{D10A} and eSpCas9(1.1)^{D10A}. Significances were calculated with two-way ANOVA followed by Šidák's test for multiple comparisons; ****P<0.0001, **0.001<P<0.01, *0.01<P<0.05; P>0.05 was considered non-significant (ns).

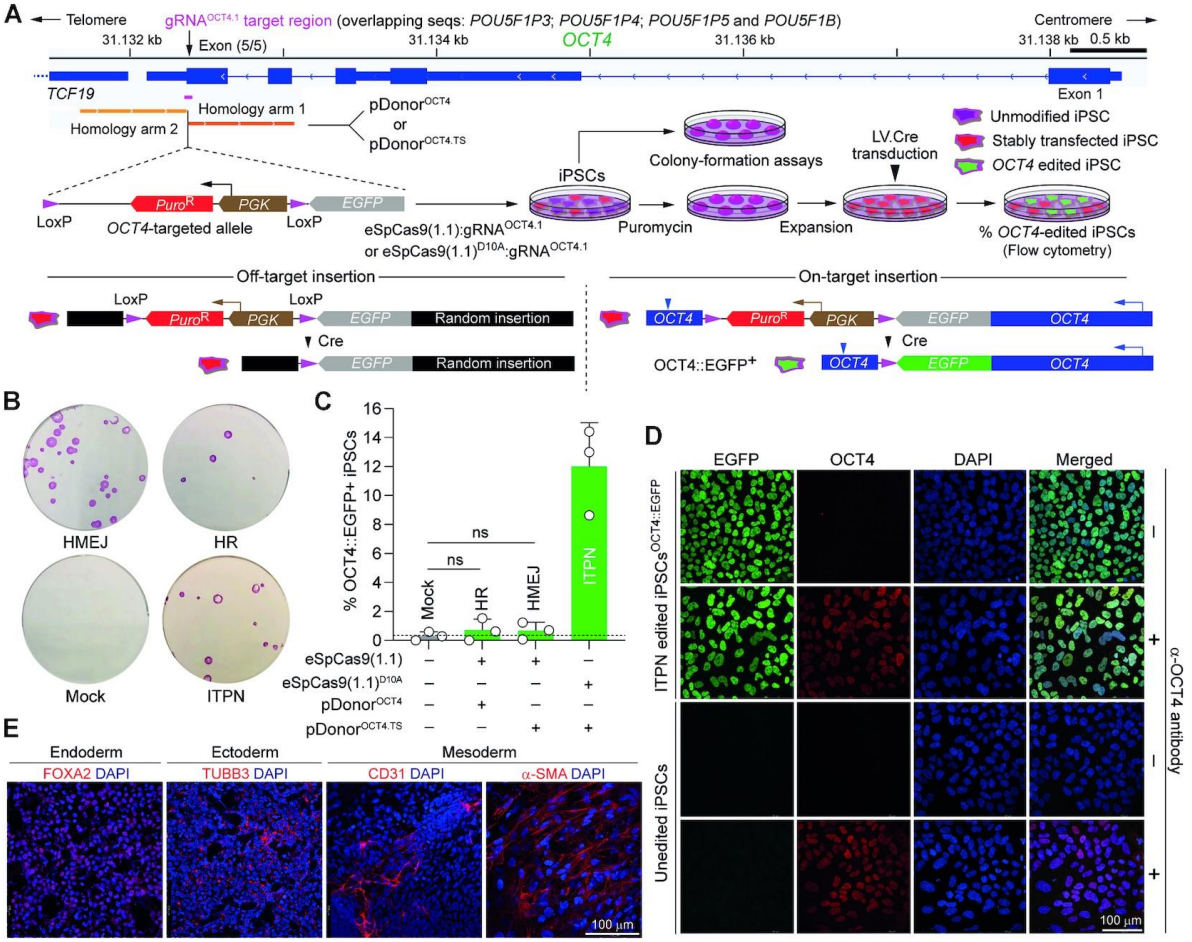


Figure 9. Testing DSB- versus SSB-dependent genome editing strategies at essential *OCT4* alleles in human iPSCs using high-specificity CRISPR complexes. (**A**) Experimental setup for tracking *OCT4* gene editing events. iPSCs exposed to the indicated reagents designed to elicit canonical HR, HMEJ or ITPN were traced by colony-formation assays upon puromycin selection and alkaline phosphatase staining and by a genetic assay reporting live-cell *OCT4* gene targeting events upon Cre recombinase delivery. (**B**) Detection of stably transfected iPSC colonies. Picture of a representative colony-formation assay is shown. (**C**) Detection of *OCT4* gene editing events. The frequencies of *OCT4* edited cells (OCT4::EGFP+) in puromycin-resistant iPSC populations were determined by EGFP-directed flow cytometry following transduction with Cre-expressing lentivector particles (20 vector particles per cell). Data are presented as mean ± S.D. of independent biological replicates (n = 3). (**D**) Confocal microscopy analysis of iPSCs edited at *OCT4* through ITPN. OCT4::EGFP-expressing iPSCs engineered through ITPN and Cre delivery (iPSC$^{OCT4::EGFP}$) were analysed through immunofluorescence microscopy for detecting OCT4 and EGFP, respectively. Nuclei were stained with DAPI. The merge of the three fluorescence signals highlights the nuclear localization of the OCT4::EGFP fusion product. Unedited iPSCs served as negative controls. iPSC and iPSC$^{OCT4::EGFP}$ specimens not incubated with the OCT4-specific primary antibody served as staining controls. (**E**) Assessing the multi-lineage differentiation capacity of iPSCs edited at OCT4 through ITPN. iPSCs$^{OCT4::EGFP}$ generated by ITPN using high-specificity eSpCas9(1.1)^{D10A} were induced to differentiate into cell lineages corresponding to the three embryonic germ layers, i.e. mesoderm, ectoderm and endoderm. Immunofluorescence microscopy detected the indicated embryonic germ layer-specific markers. Nuclei were stained with DAPI.

DISCUSSION

In this work, we have identified high-specificity SpCas9 nucleases that once combined with donor constructs tailored for HR or HMEJ can, in a locus-dependent manner, trigger genome editing to similar or higher levels than those elicited by the parental SpCas9 nuclease, i.e. SpCas9-KA (8), SpCas9-KARA (8), eSpCas9(1.1) (8), Sniper-Cas9 (11). These results contrast with those obtained with xCas9-3.7 (10), evoCas9 (9) and SpCas9-HF1 (7) in that these high-specificity nucleases normally yield the lowest frequencies of genome-edited cells. Potentially, the modulation of DNA binding, catalytic checkpoint

thresholds (4–6) and/or post-cleavage residence times (66) by different sets of SpCas9 mutations controls target-donor engagement and ultimate gene knock-in proficiencies. It is equally possible that specific chromatin contexts have a bearing on gene knock-ins involving different SpCas9 variants (67,68). Notwithstanding the individual mechanisms or combinations thereof, the genome-editing levels reached by delivering HDR-tailored donor constructs together with different SpCas9 variants largely correlate with the DNA cleaving activities of the latter tools as scored through gene knockout assays (44).

Although the mechanisms underlying recombination between target and HMEJ donors have not been dissected, it is sensible to postulate the participation of canonical HR and MMEJ factors in that HMEJ donors, similarly to HR and MMEJ donors, have long homology tracts and are substrates to DNA end-processing, respectively. Regardless, consistent with earlier investigations using parental SpCas9 (22–24), HMEJ donors were the most proficient gene knock-in substrates once combined with the above-mentioned high-efficiency SpCas9 nucleases, independently of cell type or endogenous locus targeted.

Clearly, off-target chromosomal DSBs are undesirable in that these lesions are *bona fide* substrates for NHEJ processes and, as such, they are prone to mutations and to donor DNA 'capture' at unintended genomic locations. The latter by-products arise most frequently when free-ended linear DNA substrates are presented in cell nuclei, such as those resulting from 'double-cut' donors (56). In fact, the 'capture' of free-ended double-stranded DNA at chromosomal DSBs forms the basis of pipelines for genome-wide detection of programable nuclease off-target activities (69,70). Moreover, in addition to reducing genome-editing fidelity, off-target exogenous DNA insertions heighten cellular transformation risks. Further to this point, the emergence of severe adverse events in gene therapy clinical trials using retroviral vectors harbouring strong viral enhancers offers a cautionary example of such insertional oncogenesis phenomena (71). Importantly, we have demonstrated that off-target 'capture' of exogenous DNA resulting from the processing of HMEJ donors are minimized via using high-specificity instead of parental SpCas9 nucleases. Hence, the high-specificity SpCas9 nucleases identified here as efficient tools for DSB-dependent genome editing are expected to be particularly suited for gene knock-ins entailing HMEJ and, possibly, other types of 'double-cut' donors, such as those prone to NHEJ or MMEJ.

Genetic and pharmacological approaches that, by modulating DSB repair pathway choice, favour precise HDR-mediated genome editing, are under intense investigation (72). High-specificity SpCas9 nucleases were recently shown to have potential in this regard. Specifically, systematic experiments using double-stranded oligonucleotide donors revealed that high-specificity SpCas9 variants can, in a target site-dependent manner, bias DSB repair towards HDR at the expense of non-homologous end-joining (73). In most instances, however, HDR events remain underrepresented. Contrary to DSBs, nicks are non-canonical substrates for mutagenic DNA end-joining processes. By recruiting SSB-dependent HR pathways, ITPN genome editing strategies (24,35,37,41), generically based on tandem nicking of donor and target DNA by SpCas9 nickases (4), introduce a low mutagenic burden in edited cell populations. As a result, these approaches are particularly fitting for minimizing haploinsufficiency (28), for clonal screening-free generation of model cells and organoids as well as for biallelic, multiplexing and allele-specific gene editing (24,39–41). In this study, we have identified high-specificity SpCas9^{D10A} nickases capable of eliciting ITPN genome editing to the same extent as that triggered by the parental SpCas9^{D10A} protein. Significantly, at the *CCR5* and *AAVS1* safe harbours, ITPN setups comprising members from this nickase panel (i.e. SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A} and Sniper-Cas9^{D10A}) outperformed the reference HR setup involving regular donor constructs and the SpCas9 nuclease. Importantly, indel 'footprints' installed at target and off-target sequences in genome-edited cell populations by high-specificity SpCas9^{D10A} nickases were rare and undetected, respectively. In contrast, cell populations edited through regular and high-specificity SpCas9 nucleases had over 80% of their target alleles disrupted as quantified by amplicon deep sequencing. This data underscores the high and low mutagenic burdens imposed on cells subjected to SpCas9 nucleases and nickases, respectively.

Improving the efficiency and precision of stem cell engineering is in demand owing to the increasing role that these technologies are having in science and medicine. P53-dependent cytostatic and cytotoxic responses triggered by DSBs (targeted or otherwise) limits the efficacy of genome editing in stem cells, e.g. PSCs and hematopoietic stem cells (HSCs) (29,30). To assess P53 signaling in cells

with high sensitivity to DNA damage, we exposed human iPSCs to regular and high-specificity SpCas9 nucleases, or to their respective D10A nickase counterparts, along with specific or promiscuous gRNAs. We found that in contrast to SpCas9 nucleases, neither regular nor high-specificity SpCas9^{D10A} nickases significantly activate the canonical P53 signalling pathway. As a corollary, cell therapy products derived from human iPSCs engineered with high-specificity Cas9 nickases might offer a heighten safety profile over those made through nuclease exposure. Indeed, DSB-mediated activation of signalling pathways has been shown to select for cells with potentially harmful loss-of-function or dominant-negative mutations in the tumor-suppressor P53 transcription factor or gain-of-function mutations in the K-RAS oncoprotein (31,32). Further to this point, PSCs are capable of 'spontaneously' acquiring cancer associated P53 mutations in a recurrent fashion (58). Therefore, these cells are more resistant to DSBs and, as a result, more prone to expansion than their wild-type counterparts once exposed to programmable nucleases. Moreover, recent mouse model data support the conclusion that p53 mutant cells, rather than progressing to full malignancy in a strictly haphazard fashion, suffer instead a more deterministic series of genetic instability events (74).

ITPN genome editing permits accessing in a seamless fashion challenging genomic sequences in the form of target DNA sharing high homology to off-target sites and/or coding for essential cellular functions (28). By targeting the pluripotency supporting *OCT4* gene as such a genomic locus, we provide evidence for the utility of high-specificity nicking CRISPR complexes over their DNA cleaving counterparts for achieving gene knock-ins at essential and non-unique allelic sequences in iPSCs. In this context, ITPN and complementary DSB-free technologies, such as those based on prime editors, should widen the options for precise genome editing at challenging (or otherwise) genomic sequences (75). Prime editors consist of Cas9 nickases fused to engineered reverse transcriptases and extended prime editing (PE) gRNAs (pegRNAs) that simultaneously define target and editing sequences. In contrast to ITPN and other HDR-based strategies, PE does not require delivery of donor DNA templates and allows for efficient DNA insertions of up to ~44 bp even if substantial pegRNA optimization is typically necessary (75,76). Moreover, work from our laboratory and that of others has recently disclosed that PE is more limited in non-cycling than in cycling cells (77,78). Yet, differently from HDR-based genome editing, it can perform in post-mitotic cells in vitro and in vivo (75,77). Recent developments on PE technologies that comprise the use of dual pegRNAs and site-specific recombinases permit replacing target sequences with up 250-bp of foreign DNA and inserting whole transgenes at a prime editor-placed recombination site, respectively (75). These combinatorial approaches are powerful and versatile despite requiring the delivery of large and multicomponent reagents into target cells. Moreover, PE based on dual pegRNAs is not amenable to large DNA insertions whilst, when compared to conservative HR-based ITPN, combinatorial PE and site-specific recombination is less amenable to subtle genomic edits, such as those involving endogenous gene repair, due to 'footprint' installation in the form of recombinase target sites.

In conclusion, genome editing based on high-specificity CRISPR-Cas9 complexes and donor DNA constructs prone to defined HDR processes (i.e. HR, HMEJ or ITPN) constitute a complementary set of precision genetic engineering strategies with enhanced performances and heightened safety profiles. Indeed, the HR, HMEJ and ITPN genome editing strategies investigated here can be selected based on specific experimental or biotechnological contexts and associated goals. Namely, HMEJ as the most robust strategy across different genomic target sites (**Supplementary Figure S11**) and ITPN as the least mutagenic and cytotoxic should be particularly suited for applications profiting from high-efficiency and high-fidelity genome editing, respectively (**Figure 6** and **Supplementary Figure S12**). Regarding the latter parameter, we found that SpCas9^{D10A} nickases are poor triggers of P53 signalling in human iPSCs, which makes them a fitting tool for the genomic engineering of cells with high sensitivity to DNA damage, e.g. pluripotent and tissue-specific stem cells.

DATA AVAILABILITY

All data assembled for and analyzed in this study are included in the article and additional files. The libraries of next-generation sequencing reads are deposited at the NCBI Sequence Read Archive (SRA) database under BioProject ID PRJNA879334. The raw flow cytometry datasets are deposited in the FlowRepository under repository IDs: FR-FCM-Z5P9 (Detection of *OCT4* gene editing events), FR-FCM-Z5PA (Cleaving and nicking SaCas9 mediated gene editing), FR-FCM-Z5PB (Cleaving and nicking high-specificity Cas9 variants mediated gene editing). The donor DNA constructs designed for human safe

harbor targeting through HR, HMEJ and ITPN and for expressing parental and high-specificity SpCas9^{D10A}:gRNA complexes are available through the Addgene plasmid repository. AY27_pU6.gRNA.CLYBL (#199238); AM77_pU6.Sa-gRNA.CLYBL (#199237); AZ64_pE.DonorCLYBL.TS (#199228); AD59_pEP.DonorCLYBL.TS (#199227); BB44_pmc.DonorR5.TS (#199223); AP76_pU.CAG.Cas9-D10A-K848A.rBGpA (alias SpCas9-KA-D10A) (#199253); AP70_pU.CAG.Cas9-D10A-K848A-R1060A.rGBpA (alias SpCas0-KARA-D10A) (#199254); AA69_pU.CAG.Cas9-eSp(1.1)-D10A.rBGpA.2NLS (alias eSpCas9(1.1)-D10A) (#199252); BA31_pU.CAG.SaCas9-D10A.rBGpA (alias SaCas9-D10A) (#199251); AB65_pCAG.Cas9-D10A.rBGpA (alias SpCas9-D10A) (#199256) and AW01_pU.CAG.eSpCas9(1.1).rBGpA (#199255).

SUPPLEMENTARY DATA

Supplementary Data are available online at doi: 10.1093/nar/gkad165.

ACKNOWLEDGMENTS

The authors are grateful to Aart Jochemsen (Department of Cell and Chemical Biology, Leiden University Medical Center, The Netherlands) for providing the MDM2 inhibitor Nutlin-3a, and to all laboratory members for helpful discussions.

FUNDING

Q.W. held a Ph.D. fellowship from the China Scholarship Council–Leiden University Joint Scholarship Programme. Funding for open access charge: Departmental funds.
Conflict of interest statement. None declared.

REFERENCES

1. Ernst, M.P.T., Broeders, M., Herrero-Hernandez, .P, Oussoren, E., van der Ploeg, A.T. and Pijnappel, W. (2020) Ready for Repair? Gene Editing Enters the Clinic for the Treatment of Human Disease. *Mol. Ther. Methods Clin. Dev.*, **18**, 532-557.
2. Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y. and Davis, G.D. (2017) Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.*, **8**, 14958.
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. *et al*. (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, **520**, 186-191.
4. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
5. Sternberg, S.H., LaFrance, B., Kaplan, M. and Doudna, J.A. (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature*, **527**, 110-113.
6. Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A. and Yildiz, A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.*, **3**, eaao0027.
7. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490-495.
8. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X. and Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84-88.
9. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F. et al. (2018) A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol.*, **36**, 265-271.
10. Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W.X., Chen, L.W., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z. et al. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556**, 57-63.
11. Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.H., Lee, K., Jung, I., Kim, D., Kim, S. et al. (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.,* **9**, 3048.
12. Bollen, Y., Post, J., Koo, B.K. and Snippert, H.J.G. (2018) How to create state-of-the-art genetic model systems: strategies for optimal CRISPR-mediated genome editing. *Nucleic Acids Res.*, **46**, 6435-6454.

13. Chen, X. and Gonçalves, M. (2018) DNA, RNA, and Protein Tools for Editing the Genetic Information in Human Cells. *iScience*, **6**, 247-263.
14. Jang, H.K., Song, B., Hwang, G.H. and Bae, S. (2020) Current trends in gene recovery mediated by the CRISPR-Cas system. *Exp. Mol. Med.*, **52**, 1016-1027.
15. Cho, S.W., Kim, S., Kim, J.M. and Kim, J.S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol.*, **31**, 230-232.
16. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.
17. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.
18. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
19. He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J. and Feng, B. (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.*, **44**, e85.
20. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z. et al. (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144-149.
21. Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T. et al. (2014) Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.*, **5**, 5560.
22. Yao, X., Wang, X., Hu, X., Liu, Z., Liu, J., Zhou, H., Shen, X., Wei, Y., Huang, Z., Ying, W. et al. (2017) Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res.*, **27**, 801-814.
23. Zhang, J.P., Li, X.L., Li, G.H., Chen, W., Arakaki, C., Botimer, G.D., Baylink, D., Zhang, L., Wen, W., Fu, Y.W. et al. (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.*, **18**, 35.
24. Chen, X., Janssen, J.M., Liu, J., Maggio, I., t Jong, A.E.J., Mikkers, H.M.M. and Gonçalves, M. (2017) In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun.*, **8**, 657.
25. Hustedt N. and Durocher D. (2016) The control of DNA repair by the cell cycle. *Nat. Cell Biol.*, **19**, 1-9.
26. Frock, R.L., Hu, J, Meyers, R.M., Ho, Y.J., Kii, E. and Alt, F.W. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.*, **33**, 179-186.
27. Kosicki, M., Tomberg, K. and Bradley, A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **3**, 765-71.
28. Chen, X., Tasca, F., Wang, Q., Liu, J., Janssen, J.M., Brescia, M.D., Bellin, M., Szuhai, K., Kenrick, J., Frock, R.L. et al. (2020) Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking. *Nucleic Acids Res.*, **48**, 974-995.
29. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C. et al. (2018) p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.*, **24**, 939-946.
30. Schiroli, G., Conti, A., Ferrari, S., Della Volpe, L., Jacob, A., Albano, L., Beretta, S., Calabria, A., Vavassori, V., Gasparini, P. et al. (2019) Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell*, **24**, 551-565 e558.
31. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. and Taipale, J. (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.*, **24**, 927-930.
32. Sinha, S., Barbosa, K., Cheng, K., Leiserson, M.D.M., Jain, P., Deshpande, A., Wilson, D.M., 3rd, Ryan, B.M., Luo, J., Ronai, Z.A. et al. (2021) A systematic genome-wide mapping of oncogenic mutation selection during CRISPR-Cas9 genome editing. *Nat. Commun.*, **12**, 6512.
33. Urnov, F.D. (2018) A path to efficient gene editing. *Nat. Med.*, **24**, 899-900.

34. Gonçalves M.A., van Nierop, G.P., Holkers, M. and de Vries, A.A. (2012) Concerted nicking of donor and chromosomal acceptor DNA promotes homology-directed gene targeting in human cells. *Nucleic Acids Res.*, **40**, 3443-3455.
35. Nakajima, K., Zhou, Y., Tomita, A., Hirade, Y., Gurumurthy, C.B. and Nakada, S. (2018) Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res.*, **28**, 223-230.
36. Kocher, T., Wagner, R.N., Klausegger, A., Guttmann-Gruber, C., Hainzl, S., Bauer, J.W., Reichelt, J. and Koller, U. (2019) Improved Double-Nicking Strategies for COL7A1-Editing by Homologous Recombination. *Mol. Ther. Nucleic Acids.*, **18**, 496-507.
37. Hyodo, T., Rahman, M.L., Karnan, S., Ito, T., Toyoda, A., Ota, A., Wahiduzzaman, M., Tsuzuki, S., Okada, Y., Hosokawa, Y. et al. (2020) Tandem Paired Nicking Promotes Precise Genome Editing with Scarce Interference by p53. *Cell Rep.*, **30**, 1195-1207 e1197.
38. Rahman, M.L., Hyodo, T., Karnan, S., Ota, A., Hasan, M.N., Mihara, Y., Wahiduzzaman, M., Tsuzuki, S., Hosokawa, Y. and Konishi, H. (2021) Experimental strategies to achieve efficient targeted knock-in via tandem paired nicking. *Sci. Rep.*, **11**, 22627.
39. Bollen, Y., Stelloo, E., van Leenen, P., van den Bos, M., Ponsioen, B., Lu, B., van Roosmalen, M.J., Bolhaqueiro, A.C.F., Kimberley, C., Mossner, M. et al. (2021) Reconstructing single-cell karyotype alterations in colorectal cancer identifies punctuated and gradual diversification patterns. *Nat. Genet.*, **53**, 1187-1195.
40. Fortschegger, K., Husa, A.M., Schinnerl, D., Nebral, K. and Strehl, S. (2021) Expression of RUNX1-JAK2 in Human Induced Pluripotent Stem Cell-Derived Hematopoietic Cells Activates the JAK-STAT and MYC Pathways. *Int. J. Mol. Sci.*, **22**,1.
41. Bollen, Y., Hageman, J.H., van Leenen, P., Derks, L.L.M., Ponsioen, B., Buissant des Amorie, J.R., Verlaan-Klink, I., van den Bos, M., Terstappen, L., van Boxtel, R. et al. (2022) Efficient and error-free fluorescent gene tagging in human organoids without double-strand DNA cleavage. *PLoS Biol.*, **20**, e3001527.
42. Zhang, M., D'Aniello, C., Verkerk, A.O., Wrobel, E., Frank, S., Ward-van Oostwaard, D., Piccini, I., Freund, C., Rao, J., Seebohm, G. et al. (2014) Recessive cardiac phenotypes in induced pluripotent stem cell models of Jervell and Lange-Nielsen syndrome: disease mechanisms and pharmacological rescue. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, E5383-5392.
43. Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. U. S. A*, **98**, 8241–8246.
44. Rees, H.A., Yeh, W.H. and Liu, D.R. (2019) Development of hRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.*, **10**, 2212.
45. Wang, Q., Liu, J., Janssen, J.M., Le Bouteiller, M., Frock, R.L., Gonçalves, M.A.F.V. (2021) Precise and broad scope genome editing based on high-specificity Cas9 nickases. *Nucleic Acids Res.*, **49**, 1173-1198.
46. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z. and Joung, J.K. (2015) Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.*, **33**, 1293-1298.
47. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E. et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.*, **37**, 224-226.
48. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 10-12.
49. Brinkman, E.K., Chen, T., Amendola, M. and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*., **42**, e168.
50. Papapetrou, E.P. and Schambach, A. (2016) Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol. Ther.*, **24**, 678-84.
51. Pavani, G. and Amendola, M. (2020) Targeted Gene Delivery: Where to Land. *Front Genome Ed.*, **2**, 609650.
52. Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D.F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P. et al. (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods*, **8**, 861-869.
53. Cerbini, T., Funahashi, R., Luo, Y., Liu, C., Park, K., Rao, M., Malik, N. and Zou, J. (2015) Transcription activator-like effector nuclease (TALEN)-mediated CLYBL targeting enables

enhanced transgene expression and one-step generation of dual reporter human induced pluripotent stem cell (iPSC) and neural stem cell (NSC) lines. *PLoS One*, **10**, e0116032.
54. Kulcsar, P.I., Talas, A., Huszar, K., Ligeti, Z., Toth, E., Weinhardt, N., Fodor, E. and Welker, E. (2017) Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol.*, **18**, 190.
55. Kim, S., Bae, T., Hwang, J. and Kim, J.S. (2017) Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. *Genome Biol.*, **18**, 218.
56. Holkers, M., Maggio, I., Henriques, S.F., Janssen, J.M., Cathomen, T. and Gonçalves, M.A. (2014) Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat. Methods*, **11**, 1051-1057.
57. Tasca, F., Brescia, M., Wang, Q., Liu, J., Janssen, J.M., Szuhai, K. and Gonçalves, M. (2022) Large-scale genome editing based on high-capacity adenovectors and CRISPR-Cas9 nucleases rescues full-length dystrophin synthesis in DMD muscle cells. *Nucleic Acids Res.*, **50**, 7761-7782.
58. Merkle, F.T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M. et al. (2017) Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature*, **545**, 229-233.
59. Munoz, D.M., Cassiani, P.J., Li, L., Billy, E., Korn, J.M., Jones, M.D., Golji, J., Ruddy, D.A., Yu, K., McAllister, G. et al. (2016) CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov.*, **6**, 900-913.
60. Gonçalves, E., Behan, F.M., Louzada, S., Arnol, D., Stronach, E.A., Yang, F., Yusa, K., Stegle, O., Iorio, F. and Garnett, M.J. (2019) Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss of fitness effects. *Genome Biol.*, **20**, 27.
61. Fogarty, N.M.E., McCarthy, A., Snijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D. et al. (2017) Genome editing reveals a role for OCT4 in human embryogenesis. *Nature*, **550**, 67-73.
62. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861-872.
63. Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. and Benvenisty, N. (2018) Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol.*, **20**, 610-619.
64. Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C. et al. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.*, **29**, 731-734.
65. Zhu, Z., Verma, N., Gonzalez, F., Shi, Z.D. and Huangfu, D. (2015) A CRISPR/Cas-Mediated Selection-free Knockin Strategy in Human Embryonic Stem Cells. *Stem Cell Reports*, **4**, 1103-1111.
66. Liu, S.C., Feng, Y.L., Sun, X.N., Chen, R.D., Liu, Q., Xiao, J.J., Zhang, J.N., Huang, Z.C., Xiang, J.F., Chen, G.Q. et al. (2022) Target residence of Cas9-sgRNA influences DNA double-strand break repair pathway choices in CRISPR/Cas9 genome editing. *Genome Biol.*, **23**, 165.
67. Chen, X., Liu, J., Janssen, J.M. and Gonçalves, M. (2017) The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Mol. Ther. Nucleic Acids*, **8**, 558-563.
68. Janssen, J.M., Chen, X., Liu, J. and Gonçalves, M. (2019) The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. *Mol. Ther. Nucleic Acids*, **16**, 141-154.
69. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P. et al. (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.*, **33**, 187-197.
70. Martin, F., Sanchez-Hernandez, S., Gutierrez-Guerrero, A., Pinedo-Gomez, J. and Benabdellah, K. (2016) Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *Int. J. Mol. Sci.*, **17**.
71. Fischer, A. and Hacein-Bey-Abina, S. (2020) Gene therapy for severe combined immunodeficiencies and beyond. *J. Exp. Med.*, **217**, 1.

72. Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J. and Xia, N. (2020) Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *Int. J. Mo.l Sci.*, **21**.
73. Kato-Inui, T., Takahashi, G., Hsu, S. and Miyaoka, Y. (2018) Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 with improved proof-reading enhances homology-directed repair. *Nucleic Acids Res.*, **46**, 4677-4688.
74. Baslan, T., Morris, J.P.t., Zhao, Z., Reyes, J., Ho, Y.J., Tsanov, K.M., Bermeo, J., Tian, S., Zhang, S., Askan, G. et al. (2022) Ordered and deterministic cancer genome evolution after p53 loss. *Nature*, **608**, 795-802.
75. Chen, P.J. and Liu, D.R. (2022) Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet*., doi: 10.1038/s41576-022-00541-1. Online ahead of print.
76. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. et al. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149-157.
77. Wang, Q., Liu, J., Janssen, J.M., Tasca, F., Mei, H. and Gonçalves, M. (2021) Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. *Nucleic Acids Res.*, **49**, 11986-12001.
78. Schene, I.F., Joore, I.P., Baijens, J.H.L., Stevelink, R., Kok, G., Shehata, S., Ilcken, E.F., Nieuwenhuis, E.C.M., Bolhuis, D.P., van Rees, R.C.M. et al. (2022) Mutation-specific reporter for optimization and enrichment of prime editing. *Nat. Commun.*, **13**, 1028.

Chapter 5

Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery

Qian Wang[1], Jin Liu[1], Josephine M. Janssen[1], Francesca Tasca[1], Hailiang Mei[2] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands;
[2]Department of Biomedical Data Sciences, Sequencing Analysis Support Core, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands

Nucleic Acids Research, 2021, Vol.49, No.20, ***11986–12001***
doi: 10.1093/nar/gkab938

ABSTRACT

Prime editing is a recent precision genome editing modality whose versatility offers the prospect for a wide range of applications, including the development of targeted genetic therapies. Yet, an outstanding bottleneck for its optimization and use concerns the difficulty in delivering large prime editing complexes into cells. Here, we demonstrate that packaging prime editing constructs in adenoviral capsids overcomes this constrain resulting in robust genome editing in both transformed and non-transformed human cells with up to 90% efficiencies. Using this cell cycle-independent delivery platform, we found a direct correlation between prime editing activity and cellular replication and disclose that the proportions between accurate prime editing events and unwanted byproducts can be influenced by the target-cell context. Hence, adenovector particles permit the efficacious delivery and testing of prime editing reagents in human cells independently of their transformation and replication statuses. The herein integrated gene delivery and gene editing technologies are expected to aid investigating the potential and limitations of prime editing in numerous experimental settings and, eventually, in *ex vivo* or *in vivo* therapeutic contexts.

INTRODUCTION

Programmable nucleases based on sequence-tailorable guide RNAs (gRNAs) and CRISPR-associated (Cas) nucleases are powerful genome editing tools (1,2). However, besides off-target mutagenesis (3–9), programmable nucleases often yield complex target allele disruptions and large genomic rearrangements due to double-strand break (DSB) repair by illegitimate recombination processes (10,11). Hence, recent genome editing developments include advancing from DNA cutting to DNA non-cutting technologies based on nicking Cas proteins as such (12–14), or on these RNA-programmable nickases fused to DNA modifying moieties, e.g., base editors and, more recently, prime editors (15,16). Prime editing permits installing any single base-pair substitution in addition to well-defined small insertions or deletions, while requiring neither DSBs nor donor DNA substrates (15). Prime editors consist of an extended gRNA and a $Cas9^{H840A}$ nickase fused to an engineered reverse transcriptase (RT) named, respectively, pegRNA and PE2 (**Supplementary Figure S1A**). The pegRNA is formed by a gRNA covalently linked at its 3′-end to a RT template encoding the edit-of-interest and a RT primer binding site (PBS). Site-specific genomic DNA nicking yields a 3′-ended DNA flap that, upon PBS annealing, primes RT-mediated DNA synthesis over the RNA template. After DNA copy hybridization to complementary target DNA, the edit is ultimately incorporated in the genome presumably through sequential strand-resolution reactions (**Supplementary Figure S1B**). Prime editing has two main modalities, i.e. PE2 and PE3. The former system requires the delivery of PE2:pegRNA complexes; the latter relies on the transfer of these complexes together with a conventional gRNA. In the PE3 system, gRNA-directed nicking of the non-edited DNA strand fosters the use of the edited strand as repairing template (**Supplementary Figure S1B**).

Notwithstanding their enormous potential and versatility, prime editing principles bring to the fore specific shortcomings that will need identification, careful assessment and resolution. The large size of prime editing ribonucleoprotein complexes, composed of a ∼125 nucleotide-long pegRNA and a 238-kDa fusion protein encoded by a 6.3-kb ORF, poses substantial production and delivery issues. Indeed, producing proteins >100 kDa in sufficient quantities is particularly challenging. Moreover, although viral vectors are amongst the most efficient genome-editing tool delivery systems (17), the most commonly used platform, based on ∼15 nm adeno-associated viral (AAV) particles, is unsuitable for transferring full-length prime editing sequences due to its limited packaging capacity (∼4.7-kb) (17).

Fully viral gene-deleted adenoviral vectors (also called high-capacity adenoviral vectors), hereafter named adenovector particles (AdVPs), aggregate a valuable set of characteristics, namely; (i) large packaging capacity (i.e. up to 36-kb), (ii) strict episomal nature, (iii) high genetic stability; (iv) facile cell-tropism modification and (v) efficient transduction of dividing and quiescent cells (17–21). Here, we investigate the feasibility and utility of tailoring these ∼90 nm biological nanoparticles for all-in-one transfer of full-length prime editing components and, as the cellular processes underlying or influencing prime editing outcomes are essentially unknown, exploit the latter characteristic to study the role of cell cycling on this site-specific DNA modifying principle.

MATERIALS AND METHODS

Cells

The human cervix carcinoma (HeLa) cells and the human embryonic kidney 293T (HEK293T) cells (both from American Type Culture Collection) were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific; Cat. No.: 41966-029) containing 5% and 10% fetal bovine serum (FBS; Biowest; Cat. No.: S1860-500), respectively. PEC3.30 cells were maintained in high-glucose DMEM supplemented with 10% FBS, 10 mM $MgCl_2$ and 0.4 μg ml^{-1} puromycin (Thermo Fisher Scientific; Cat. No.: A11138-03) (22). The generation and characterization of HEK293T.EGFP$^+$ reporter cells harboring a transcriptionally active EGFP allele in the presence of doxycycline, were detailed elsewhere (23). These cells were cultured in high-glucose DMEM containing 10% FBS and 200 ng ml^{-1} doxycycline. The bone marrow-derived primary human mesenchymal stem cells (hMSCs) were kept in Minimum Essential Medium α (MEM-α) (Thermo Fisher Scientific; Cat. No.: 22561-021) supplemented with 10% FBS, 100 U ml^{-1} penicillin/streptomycin (Thermo Fisher Scientific; Cat. No.: 15140–122), 1× non-essential amino acids (NEAA; Thermo Fisher Scientific; Cat. No.: 11140–050) and 1× GlutaMax (Thermo Fisher Scientific; Cat. No.: 35050-061) (24,25). The hMSCs were passaged every 3–4 days at low split ratios in culture vessels coated with 0.1% gelatin (Sigma-Aldrich; Cat. No.: G1393) for at least 2 h at 37°C. Collection of human primary cells from bone marrow was carried out from anonymous 'left-over' surgery material in accordance with the Best Practices Code of the Dutch Federation of Biomedical Scientific Societies. No informed consent is required for the use of anonymous and non-traceable body materials and the institutional ethics committee of the LUMC waived the need for donor consent. The human myoblasts derived from an healthy donor and the retinal pigment epithelial cells expressing Fucci reporters (RPE-Fucci) (26), have both been described before (27–29). The former cells were cultured in Skeletal Muscle Cell Growth Medium (Ready-to-use, PromoCell; Cat. No.: C-23060) containing 20% FBS, 100 U ml^{-1} penicillin/streptomycin and 1× GlutaMax; the latter cells were maintained in DMEM/Nutrient Mixture F-12 (DMEM/F-12; Thermo Fisher Scientific; Cat. No.: 31331-028) supplemented with 10% FBS. All cells were kept in a humidified-air 5% CO_2 atmosphere and were routinely tested for verifying the absence of mycoplasma.

Recombinant DNA

Standard recombinant DNA techniques were applied for the generation of the various constructs (Supplementary Information). The cloning schemes, annotated maps and nucleotide sequences of BF18_pLV.PURO.pegRNA^{HBB}, BF19_pLV.PURO.pegRNA.gRNA^{HBB}, BF20_pLV.BSD-EGFP.pegRNA^{EGFP}, BF21_pLV.BSD-EGFP.pegRNA.gRNA^{EGFP}, BF46_pLV.PURO.pegRNA.gRNA^{EGFP}, BF22_pUCBM21.U6.gRNA^{HBB}, BF23_pUCBM21.U6.gRNA^{EGFP}, S75_pAdVP.PE2 and S89_pAdVP.PE2.pegRNA^{CTT} are available in pages 1–35 of the **Supplementary Information**. The oligonucleotides used for assembling the various gRNA and pegRNA expression constructs are indicated in **Supplementary Table S1**.

AdVP production, purification and characterization

The production of AdVP.PE2 and AdVP.PE2.pegRNA^{CTT} was done as follows. The bacteriophage P1 Cre recombinase- and adenovirus type 5 E1-expressing PEC3.30 cells were seeded at a density of 1.8 × 10^6 cells per well of 6-well plates (Greiner Bio-One). Transfection of the AdVP molecular clones was performed 16–18 h later with the aid of 25 kDa linear polyethyleneimine (PEI, Polysciences). In brief, 6.25 μg of MssI-linearized plasmids S75_pAdVP.PE2 and S89_pAdVP.PE2.pegRNA^{CTT} were diluted in a total volume of 200 μl of 150 mM NaCl and, after adding 20.6 μl of a 1 mg ml^{-1} PEI solution (pH 7.4) to each of the DNA transfection reactions, vigorous mixing in a vortex for about 10 s ensued. The DNA-PEI complexes, assembled after a 15-min incubation period at room temperature (RT), were directly added to the medium of the producer cells. Six hours later, transfection media were substituted by fresh medium containing E1-deleted helper AdV vector AdV.SRα.LacZ.1.50 (30) at an MOI of 10 infectious units (IUs) per cell. The helper vector contains its packaging elements flanked by a direct repeat of Cre recombinase loxP recognition sites (floxed.Ψ) and is used for supporting the amplification and selective packaging of AdVP genomes (Supplementary Figure S2). Typically, AdVP producer cells express Cre recombinase and E1 peptides (18–20). In addition to these proteins, PEC3.30 cells also express a thermosensitive form of the adenoviral DNA-binding protein (DBP) that remains inactive during regular culturing at 39°C (30,31). Hence, after helper addition, PEC3.30 cells were transferred from 39°C to 34°C for extra adenoviral protein complementation in the form of properly folded thermosensitive DBP. Upon helper-triggered emergence of complete cytopathic effect (CPE), the producer cells were harvested and subjected to three cycles of freezing and thawing in liquid N_2 and 37°C water baths, respectively. Cellular debris were subsequently removed by centrifugation for 10 min at 2000 × g. The

vector particles present in the collected supernatants were then amplified via four rounds of propagation in producer cells transduced with helper AdV.SRα.LacZ.1.50 (**Supplementary Figure S2**). The fourth propagation step involved twenty T175-cm^2 culture flasks (Greiner Bio-One) each containing 27 × 10^6 producer cells. The resulting AdVPs were purified by sequential block and continuous CsCl buoyant density ultracentrifugation steps and were de-salted by ultrafiltration through Amicon Ultra-15 100K MWCO filters (MerckMillipore; Cat. No.: UFC910024) (**Supplementary Figure S3**). With the exception of the use of a AdVP molecular clone expressing mCherry from the hybrid CAG promoter (**Supplementary Figure S4**), the production of the reporter vector AdVP.mCherry followed essentially the same methodologies applied for the production of AdVP.PE2 and AdVP.PE2.pegRNA^{CTT}. The titers of purified AdVP stocks were determined via previously detailed procedures using the Quant-iT™ PicoGreen™ dsDNA Assay Kit reagents and protocol (Thermo Fisher Scientific; Cat.No.: P11496A) (30,32). Additionally, physical and transducing titers of prime editor-encoding AdVPs were also determined through qPCR assays. Quantification of physical titers was initiated by using the DNeasy Blood & Tissue kit (QIAGEN; Cat. No.: 69506) to isolate vector DNA from purified AdVP stocks. Next, six serial 3-fold dilutions of the extracted vector genomes were prepared for qPCR with iQ™ SYBR® Green Supermix (Bio-Rad; Cat. No.: L010171C) and the primers listed in **Supplementary Table S2**. Quantification of transducing titers was initiated by plating HeLa cells at a density of 8 × 10^4 cells per well of 24-well plates (Greinder Bio-One). The next day, the cells were transduced with six serial 3-fold dilutions of each purified AdVP preparation. After approximately 24 h, the transduced cells were harvest for DNA isolation by using the DNeasy Blood & Tissue kit (QIAGEN; Cat. No.: 69506). The resulting DNA was then used for qPCR quantification of transducing vector genome copies. In parallel, a standard curve was generated by using as qPCR template, eight serial 10-fold dilutions of a linearized target DNA-containing plasmid stock containing 1 × 10^7 GC μl^{-1}. The primers, cycling conditions and components of qPCR mixtures are specified in **Supplementary Tables S2** and **S3**. Data analysis was performed by using Bio-Rad CFX Manager 3.1 software and the titers were calculated based on the Ct values of standard curve and sample dilutions. The AdVP titers obtained via the different titration methods are listed in **Supplementary Table S4**. The AdVP MOI indicated in this study were based on packaged vector genome copies (GC) determined by the PicoGreen™ titration method (30,32). The structural integrity of vector genomes packaged in purified adenoviral capsids (**Supplementary Figure S3**) was assessed essentially as indicated elsewhere (30,32). In brief, 50 μl of purified AdVP stocks were treated with 8 μl of 10 mg ml^{-1} DNaseI (Sigma-Aldrich; Cat. No.: 10104159001) at 37°C for 30 min. Next, the DNaseI enzyme was inactivated by adding 2.4 μl of 0.5 M ethylenedinitrilotetraacetic acid (EDTA) solution (pH 8.0), 6 μl of 10% (w/v) sodium dodecyl sulphate (SDS) and 1.5 μl of 20 mg ml^{-1} proteinase K (Thermo Fisher Scientific; Cat. No.: EO0491), and incubating the resulting mixture at 55°C for 1 h. Vector DNA isolation was then done by using the QIAEX II Gel Extraction Kit (QIAGEN; Cat. No.: 20021) following the manufacturer's instructions. Finally, the isolated vector genomes were subjected to restriction enzyme fragment analysis (RFLA) by using the Gel-Doc XR + system and the ImageLab 4.1 software (both from Bio-Rad). Parental plasmids pAdVP.PE2 and pAdVP.PE2.pegRNA^{CTT}, digested with the same restriction enzymes applied to vector genomes, served as molecular weight references. The in silico restriction patterns corresponding to intact plasmid and vector DNA were made with the aid of SnapGene (version 5.2.4) software (**Supplementary Figure S3**).

Transduction experiments

HeLa cells, hMSCs and HEK293T cells (both regular and genetically-modified through lentiviral vector transduction) were seeded at densities of 5 × 10^4, 1.5 × 10^5, and 7 × 10^4 cells per well of 24-well plates (Greinder Bio-One), respectively. After overnight incubations, the cells were either mock-transduced or were transduced with the various AdVPs at the MOIs specified in the figures and respective legends. The cells were harvested at 2 days post-transduction for western blotting and immunofluorescence microscopy analyses and at 3 days post-transduction for prime editing analyses. For the latter analyses, genomic DNA was extracted by using the DNeasy Blood & Tissue kit (QIAGEN; Cat. No.: 69506) following the manufacturer's recommendations. Transduction efficiencies were determined by reporter-directed flow cytometry and direct fluorescence microscopy analyses of parallel cell cultures exposed to AdVP.mCherry. To compare prime editing activities mediated by AdVP transduction versus plasmid transfection, 1.5 × 10^5 HEK293T cells were seeded in wells of 24-well plates (Greinder Bio-One). The next day, plasmids encoding prime editing complexes were diluted in 150 mM NaCl to which 3.95 μl of 1 mg ml^{-1} of 25-kDa linear polyethyleneimine (pH 7.4; Polysciences) were added (**Supplementary Table S5**). The resulting transfection reactions were vortexed vigorously for about 10 sec and were

then incubated for 15 min at RT. Afterwards, the reactions were directly added to the cell media and, after 6 h, the transfection media were substituted by regular culture media. In parallel, HEK293T cells were transduced with AdVPs at the MOIs specified in the Supplementary Table S5. Three days later, genomic DNA from transfected and transduced cells were harvested using the DNeasy Blood & Tissue kit reagents and protocol (QIAGEN; Cat. No.: 69506) for prime-editing activity analysis. The transduction of dividing myoblasts and non-dividing myotubes was done as follows. Approximately 16–18 h prior to transduction, 5×10^4 human myoblasts were seeded in wells of 24-well plates and, the next day, these cells were exposed to AdVPs at the MOIs indicated in the respective figures. In parallel, 2×10^5 human myoblasts were seeded in the wells of 24-well plates pre-coated with a 0.1% (w/v) gelatin solution. Upon myoblast confluency, differentiation was triggered by switching regular culture medium for mitogen-poor differentiation medium whose composition is indicated below. Two days after differentiation initiation, the cells were treated with AdVPs at the MOIs specified in the corresponding figures. Myoblasts and myotubes were harvested for western blotting and genomic DNA extraction at 2 and 3 days post-transduction, respectively. RPE-Fucci cells were seeded at a density of 2×10^5 cells per well of six-well plates. For generating cultures containing different proportions of cycling and non-cycling cells, RPE-Fucci cells were treated at ~17 h after seeding with DMEM/F-12 medium lacking or containing different FBS concentrations (i.e. 0.1, 0.5, 1, 5 and 10%). Twelve hours later, the RPE-Fucci cells were exposed for an additional 12-h period to 10^4 GC cell^{-1} of AdVP.PE2.pegRNA^{CTT} in medium containing the various FBS amounts. Finally, the cells were harvested for western blotting and genomic DNA extraction at 60 h post-transduction. At 12, 24 and 48 h after the initiation of the various FBS treatments, the frequencies of RPE-Fucci cells at different phases of the cell cycle were determined by flow cytometry. The transduction of HEK293T.EGFP$^+$ cells stably expressing *EGFP*-specific pegRNA and gRNA was done as follows. The cells were seeded at 2×10^5 cells per well in 24-well plates and, after overnight incubation, the cells were exposed to different AdVP.PE2 MOI for 2 days. Quantification of EBFP-positive and EGFP-negative cells was performed at 7 days post-transduction via reporter-directed flow cytometry.

Cell differentiation assays

The capacity of hMSCs to differentiate into osteoblasts and adipocytes and of human myoblasts to differentiate into syncytial myotubes, was assessed as follows. To induce osteogenic differentiation, mock- and AdVP-transduced hMSCs initially seeded at a density of 1×10^4 cells per well of 48-well plates (Greiner Bio-One), were incubated for 2 weeks in osteogenic differentiation medium consisting of MEM-α supplemented with 10% FBS, 100 U ml^{-1} penicillin/streptomycin, 1× NEAA, 1× GlutaMax, 0.2 mM L-ascorbic acid 2-phosphate (Sigma-Aldrich; Cat. No.: A8960), 10 mM β-glycerophosphate (Sigma-Aldrich; Cat. No.: G6251), 2 μM dexamethasone (Sigma-Aldrich; Cat. No.: D4902) and 100 ng ml^{-1} of recombinant human bone morphogenetic protein 6 (BMP6; PeproTech; Cat. No.: 120-06). The differentiation medium was replenished every 3–4 days. Alizarin Red S staining was carried out for the detection of calcium deposits. In brief, cells were fixed with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS; pH 7.4) for at least 10 min and were then incubated at RT in the dark for 5 min in a 2% Alizarin Red S (pH 4.25) solution (Sigma-Aldrich; Cat. No.: A5533). After several washes with PBS, the treated cultures were photographed and were subsequently incubated for 30 min at RT, while shaking, in 100 μl of a 10% (v/v) acetic acid solution (Sigma-Aldrich; Cat. No.: 64-19-7) for Alizarin Red S extraction. Next, the samples were heated for 10 min at 85°C and then centrifuged for 15 min at 20 000 × g. The pH of the resulting supernatants was adjusted to 4.1–4.5 with a 32% (w/v) ammonium solution (Merck Millipore; Cat. No.: 105426) and measurements of the absorbance at OD_{405} nm were done by using a multimode plate reader (PerkinElmer VICTOR™ X3). To trigger adipogenic differentiation, mock- and AdVP-transduced hMSCs, initially seeded at a density of 2×10^4 cells per well of 24-well plates (Greiner Bio-One), were treated with 100 ng ml^{-1} of recombinant human bone morphogenetic protein 7 (BMP7; PeproTech; Cat. No.: 120-03) until reaching confluence in 3 days. Next, the cells were incubated for 3 weeks in adipogenic differentiation medium consisting of DMEM supplemented with 10% FBS, 100 U ml^{-1} penicillin/streptomycin, 50 μM Indomethacin (Sigma-Aldrich; Cat. No.: I7378), 0.25 μM dexamethasone, 0.5 mM 3-isobutyl-1-methylxanthine (IBMX; Sigma-Aldrich; Cat. No.: I5879) and 1.6 μM bovine insulin (Sigma-Aldrich; Cat. No.: I5500). The differentiation medium was replenished every 4 days. Oil Red O staining was performed for the detection of lipid droplets. In brief, the cells were fixed with 4% PFA in PBS (pH 7.4) for ~1 h and, after two washes with water and a 5-min treatment with 60% 2-propanol, they were incubated for 10 min at RT in a solution of 1.8 mg ml^{-1} Oil Red O (Sigma-Aldrich; Cat. No.: O-0625) in 60% 2-propanol. After several washes with water, the

treated cultures were photographed and were subsequently incubated for 10 min at RT, while shaking, in 2-propanol for Oil Red O extraction. Measurements of the absorbance at OD_{490} nm were carried out with the aid of a multimode plate reader (PerkinElmer VICTOR™ X3). To induce myogenic differentiation, confluent human myoblasts initially seeded at a density of 2×10^5 cells per well of 24-well plates pre-coated with 0.1% gelatin, were exposed to differentiation medium consisting of phenol red-free DMEM (Thermo Fisher Scientific; Cat. No.: 11880-028) supplemented with 100 U ml^{-1} penicillin/streptomycin, 100 μg ml^{-1} human holo-transferrin (Sigma-Aldrich; Cat. No.: T0665) and 10 μg ml^{-1} human insulin (Sigma-Aldrich; Cat. No.: I9278). Post-mitotic myotubes were detected by immunofluorescence staining with an antibody specific for the late muscle marker, sarcomeric α-actinin (**Supplementary Table S6**).

Reverse transcription-quantitative PCR (RT-qPCR)

RT-qPCR was applied for quantifying the differentiation abilities of mock- and AdVP-transduced hMSCs. Total RNA from cultures containing undifferentiated and differentiated hMSCs (osteoblasts and adipocytes) was extracted by using the NucleoSpin RNA kit according to the manufacturer's instructions (Macherey Nagel; Cat. No.: 740955). Next, reverse transcription was carried out with the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific; Cat. No.: K1691). In brief, 450 ng of RNA was incubated with 0.5 μl of 100 μM random hexamer primers and 0.5 μl of 100 μM $Oligo(dT)_{18}$ primers in 12-μl reaction volumes at 65°C for 5 min followed by an incubation at 4°C for 2 min. After a brief spinning, the mixtures were immediately chilled on ice and reverse transcription components consisting of 1 μl of 20 U μl^{-1} RiboLock RNase Inhibitor, 1 μl of 200 U μl^{-1} RevertAid H Minus M-MuLV Reverse Transcriptase, 2 μl of 10 mM dNTP Mix and 4 μl of 5× Reaction Buffer, were directly added to each sample. Next, the samples were incubated at 25°C for 5 min followed by an incubation at 42°C for 1 h. Finally, the reactions were terminated by heating the samples at 70°C for 5 min. The resulting cDNA templates were then diluted 4-fold in nuclease-free water and 1 μl of diluted cDNA was subjected to qPCR by using the iQ™ SYBR® Green Supermix (Bio-Rad; Cat. No.: L010171C) and the primers listed in **Supplementary Table S2**. *GAPDH* transcripts served as RT-qPCR targets for gene expression normalization. The signal outputs were detected by using the CFX Connect Real-Time PCR Detection System (Bio-Rad) and the relative expression of each target gene was analyzed through the $2^{-\Delta\Delta Ct}$ method. The qPCR cycling conditions and mixture components are specified in **Supplementary Tables S2** and **S3**, respectively.

Western blotting

Laemmli buffer consisting of 8.0% glycerol, 3% SDS and 200 mM Tris–HCl (pH 6.8) was applied for lysing cells for 5 min at 100°C. Protein concentrations in the resulting cell lysates were determined by using a DC™ protein assay kit (Bio-Rad; Cat. No.: 5000111) according to the manufacturer's instructions. Proteins were loaded in equal amounts and were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Afterwards, the resolved proteins were transferred onto 45-μm polyvinylidene difluoride (PVDF) membranes (Merck Millipore; Cat. No.: IPVH00010) that were subsequently blocked with 5% (w/v) non-fat dry milk (Campina Elk; Cat. No.: 112349) dissolved in Tris-buffered saline (TBS; 50 mM Tris–HCl pH 7.6; 150 mM NaCl) with 0.1% (v/v) Tween 20 (TBST) at RT for 1 h. After the blocking step, the membranes were incubated overnight at 4°C with primary antibodies raised against S. pyogenes Cas9 (Abcam; Cat. No.: ab191468; 1:1000 dilution), myosin heavy chain (Sigma-Aldrich; Cat. No.: M4276; 1:500 dilution), sarcomeric α-actinin (Sigma-Aldrich; Cat. No.: A7811; 1:1000 dilution), dystrophin (Abcam; Cat. No.:ab15277; 1:500 dilution), ki-67 (Sigma-Aldrich; Cat. No.: AB9260; 1:1000 dilution), α/β-tubulin (Cell signalling; Cat. No.: 2418S; 1:1000 dilution) and vinculin (Sigma-Aldrich; Cat. No.: V9131; 1:1000 dilution) all diluted in TBST supplemented with 5% bovine serum albumin (BSA). Next, the membranes were washed with TBST thrice and the antigens were probed at RT for 2 h with horseradish peroxidase (HRP)-conjugated secondary antibodies raised against mouse IgG (Sigma-Aldrich; Cat. No.: NA931V; 1:5000 dilution) or rabbit IgG (Cell signalling; Cat. No.: 7074S; 1:1000 dilution) diluted in TBST containing 1% (w/v) non-fat dry milk. Target proteins were detected with the Clarity™ Western ECL Substrate (Bio-Rad; Cat. No.: 1705060) and the ChemiDoc Imaging System (Bio-Rad; Cat. No.: 17001402).

Flow cytometry

Flow cytometry was applied for quantifying cell transduction efficiencies and RPE-Fucci cell fractions at different stages of the cell cycle. In brief, cells were washed with PBS (pH 7.4) and, after trypsin treatment and centrifugation, collected cells were resuspended in PBS containing 0.5% BSA and 2 mM EDTA (pH 8.0). Flow cytometry was performed in a BD LSR II flow cytometer (BD Biosciences) using

mock-transduced cells as controls for establishing background fluorescence thresholds. At least 10,000 viable single cells were acquired per sample. Data were analysed with the aid of FlowJo 10.5.0 software (Tree Star).

Direct fluorescence microscopy

Transduction of hMSCs and human myoblasts was monitored by direct fluorescence microscopy. The nuclei were stained with 10 µg ml^{-1} Hoechst 33342 (Thermo Fisher Scientific; Cat. No.: H3570) for 10 min. Subsequently, the cells were washed thrice with PBS, after which, regular culture medium was added to the cell cultures. The mCherry- and Hoechst 33342-specific signals were detected by using an inverted DMI8 fluorescence microscope equipped with a DFC450C camera and acquired images were examined with the LAS X software (Leica Microsystems).

Confocal immunofluorescence microscopy

Cells were fixed with 4% PFA in PBS for 10 min at RT and then washed thrice with PBS before being permeabilized in 0.5% Triton X-100 in TBS (pH 7.6) at RT for 10 min. After three 10-min washes with 0.1% Triton X-100 in TBST, a blocking step was performed by incubating the permeabilized cells in a blocking solution consisting of TBS, 0.1% Triton X-100, 2% BSA and 0.1% sodium azide for 1 h at RT. Next, the cells were incubated overnight at 4°C with the appropriate primary antibodies diluted in blocking solution (**Supplementary Table S6**). The specimens were subsequently subjected to three 10-min washes with TBST and the target antigens were probed with fluorophore-conjugated secondary antibodies diluted in blocking solution for 1 h in the dark at RT (**Supplementary Table S6**). Finally, after three 10-min washes with TBST, ProLong Gold Antifade Mounting reagent containing DAPI (Thermo Fisher Scientific; Cat. No.: P36931) was used for mounting the specimens. Immunofluorescence microscopy images were acquired with the aid of an upright Leica SP8 confocal microscope equipped with Leica hybrid detectors HyD and were analysed with the LAS X software (Leica Microsystems).

Edu labelling

The 5-ethynyl-2′-deoxyuridine (EdU) labelling of myoblasts and myotubes was carried out by using the Click-iT® EdU Flow Cytometry Assay Kit (Thermo Fisher Scientific; Cat. No.: C10425). In brief, myoblasts and myotubes were treated with 10 µM EdU solution for 2 h at 37°C and were subsequently fixed with 4% PFA at RT for 15 min. After two washes with 3% BSA in PBS, a permeabilization step was performed by incubating the fixed cells in 0.5% Triton X-100 in TBS for 20 min at RT. Afterwards, the permeabilized cells were washed twice with 3% BSA in PBS and were incubated for 30 min in the dark at RT with 500 µl of Click-iT® reaction cocktail consisting of 427.5 µl of TBS, 20 µl of $CuSO_4$, 2.5 µl of Alexa Fluor® 488 azide and 50 µl of 1× Reaction buffer additive. Next, a blocking TBS solution containing 0.1% Triton X-100, 2% BSA and 0.1% sodium azide was applied for 1 h at RT after which the appropriate primary antibodies diluted in blocking solution were added. After overnight incubation at 4°C in the dark, the specimens were washed thrice for 10 min with TBST and fluorophore-conjugated secondary antibodies diluted in blocking solution were added for 1 h at RT in the dark (**Supplementary Table S6**). Finally, ProLong Gold Antifade Mounting reagent containing DAPI (Thermo Fisher Scientific; Cat. No.: P36931) was used for mounting the specimens. Immunofluorescence microscopy images were acquired with the aid of an inverted DMI8 fluorescence microscope equipped with a DFC450C camera and with an upright Leica SP8 confocal microscope equipped with Leica hybrid detectors HyD. The resulting micrographs were analysed with LAS X software (Leica Microsystems).

DNA content analysis

Staining with the DNA dye Hoechst 33342 was performed to determine RPE-Fucci cell fractions at different stages of the cell cycle based on their DNA content. In brief, 2×10^5 cells RPE-Fucci cells were seeded in wells of 6-well plates. After overnight incubation, the cells were treated with DMEM/F-12 medium lacking or containing FBS (i.e. 0.1, 0.5, 1, 5 and 10% FBS). At 12 h, 24 h and 48 h after the initiation of the various FBS treatments, the cells were stained with 2 µl of a 10 mg ml^{-1} Hoechst 33342 solution (Thermo Fisher Scientific; Cat. No.: H3570) for 10 min at 37°C. After several washes with PBS, the DNA content profiles in the various cell populations were determined by using a BD LSR II flow cytometer (BD Biosciences).

Target site genotyping assays

Prime editing activities were assessed by DNA sequencing assays. The hMSCs, human myoblasts and RPE-Fucci cells were exposed to AdVP.PE2.pegRNA^{CTT} at the MOIs indicated in the respective figures. At 3 days post-transduction, genomic DNA was extracted via the DNeasy Blood & Tissue Kit reagents and protocol. The 273-bp target-specific PCR product was amplified with Phusion High-Fidelity Polymerase (Thermo Fisher Scientific; Cat. No.: #F-530L) and the primers with adapter tag overhangs listed in **Supplementary Table S7**. The cycling conditions and PCR mixtures used are specified in **Supplementary Tables S8** and **S9**, respectively. The resulting amplicons were purified with AMPure XP beads (Beckman Coulter; Cat. No.: A63881) and were subsequently subjected to PCR barcoding using Illumina tag-specific primer pairs with unique sequence combinations for demultiplexing and sample identification (**Supplementary Table S10**). The cycling parameters and PCR mixtures used for the preparation of barcoded amplicons are indicated in **Supplementary Tables S8** and **S11**, respectively. After purification using AMPure XP beads, the concentrations of barcoded amplicons were determined by using the Qubit dsDNA HS assay kit (Thermo Fisher Scientific; Cat. No.: Q32854) with the Qubit2.0 fluorometer. Finally, purified amplicons were pooled together in an equal molar ratio and then were subjected to Illumina MiSeq deep sequencing for obtaining 100, 000 paired-end reads. The analyses of deep sequencing data were carried out with the aid of CRISPResso2 software (33) after demultiplexing and adapter trimming of the paired-end MiSeq raw reads (R1 and R2 fastq files) with Cutadapt version 2.10 (34). CRISPResso2 was run in standard prime editing mode with multiple alleles option for quantifying the frequencies of intended prime-editing events and imprecise byproducts consisting of scaffold-derived sequences and indels. Prime editing activities in heterozygous hMSCs were calculated as: % (number of reads with the desired CTT insertion that do not contain indels)/(number of total aligned reads). The prime-edited reads in homozygous myoblasts, myotubes and RPE-Fucci cells were divided into three sub-types: composite CTT plus G edits, partial CTT edits and partial G edits. The frequencies of each individual outcome was calculated as: % (number of reads with CTT + G insertion that do not contain indels)/(number of total aligned reads); % (number of reads with CTT insertion that do not contain indels)/(number of total aligned reads); and % (number of reads with G mutation that do not contain indels)/(number of total aligned reads), respectively. Indel yields in all the experiments were calculated as: % (number of reads with indels that do not contain scaffold incorporated events)/(number of total aligned reads). The codes applied in each round of the CRISPResso2 analyses are available as Supplementary Information. Prime editing activities were also assessed through the analysis of Sanger sequencing chromatogram peaks using the EditR method (35). In brief, genomic DNA from AdVP-transduced cells was extracted at 3 days post-transduction with the DNeasy Blood & Tissue kit following the manufacturer's recommendations. Next, the target sites were amplified with GoTaq G2 Flexi DNA Polymerase (Promega; Cat. No.: M7805). The primer sequences, PCR mixture compositions and cycling conditions are specified in **Supplementary Tables S12** and **S13**. Afterwards, the resulting amplicons were purified with the QIAEX II Gel Extraction Kit protocol and subjected to automated Sanger sequencing.

Lentiviral vector production and purification

The production of lentiviral vectors encoding pegRNAs and pegRNA/gRNA pairs was essentially carried out as follows. Twenty million HEK293T cells were seeded per 175-cm^2 culture flask (Greiner Bio-One). The next day, 30-µg plasmid mixtures were diluted in 150 mM NaCl to a final volume of 1 ml. The plasmid mixtures consisted of a 2:1:1 molar ratio of each lentiviral transfer vector, the packaging construct psPAX2 (Addgene #12260) and the vesicular stomatitis virus glycoprotein-G-pseudotyping construct pLP/VSVG (Invitrogen). The lentiviral transfer vectors used were BF18_pLV.PURO.pegRNA^{HBB}, BF19_pLV.PURO.pegRNA.gRNA^{HBB}, BF20_pLV.BSD-EGFP.pegRNA^{EGFP}, BF21_pLV.BSD-EGFP.pegRNA.gRNA^{EGFP} and BF46_pLV.PURO.pegRNA.gRNA^{EGFP} (**Supplementary Information**). In parallel, 99 µl of a 1 mg ml^{-1} PEI solution was diluted in 150 mM NaCl to a final volume of 1 ml. This PEI solution was subsequently added to each of the plasmid mixtures and, after vigorous homogenization in a vortex for approximately 10 s, a 15-min incubation period at RT ensued. Subsequently, the assembled PEI-DNA complexes were directly added to the medium of the HEK293T producer cells. After overnight incubation, the transfection media were replaced by 15 ml of regular culture medium. At 3 days post-transfection, the supernatants containing lentiviral vector particles were collected. Cellular debris were removed by centrifugation and filtration of the supernatants through 0.45-µm pore-sized HT Tuffryn membrane filters (Pall Life Sciences; Cat. No.: PN4184). The resulting clarified supernatants were then gently added onto 5-ml 20% (w/v) sucrose cushions in 35.8-ml polyallomer tubes (Beckman Coulter; Cat. No.: 326823). After ultracentrifugation (30 000 RPM for 2 h at 4°C) in an

Optima LE-80K centrifuge (Beckman Coulter) using the SW32Ti rotor, pelleted vector particles were resuspended overnight at 4°C in 400 µl of ice-cold PBS supplemented with 1% BSA. The titers of the purified lentiviral vector stocks were determined by converting 1 ng of p24 antigen to 2500 lentiviral vector transducing units (36,37) after applying the materials and protocol of the RETROTEK HIV-1 p24 antigen ELISA kit (ZeptoMetrix, Cat. No.: 0801111).

Generation of cells stably expressing prime editing RNAs

The generation of HEK293T cells stably expressing HBB-specific and EGFP-specific pegRNAs and gRNAs was initiated by seeding HEK293T cells in regular growth medium at a density of 5×10^4 cells per well of 24-well plates. At ~16–24 h after seeding, the cells were exposed to medium containing 8 µg ml^{-1} polybrene and lentiviral vectors at the indicated MOIs. After overnight incubation, the inocula were substituted by fresh culture medium. At 48 h post-transduction, the cells were transferred to 6-well plates containing regular growth medium supplemented with 2 µg ml^{-1} puromycin (Thermo Fisher Scientific; Cat. No.: A11138-03) or 5 µg ml^{-1} blasticidin (Thermo Fisher Scientific; Cat. No.: R21001). Parental mock-transduced cells served as negative controls during the drug selection procedure.

Statistical Analyses

Statistical analyses were performed with the aid of GraphPad Prism software (version 8.0.1) on datasets derived from independent biological replicates. Statistical significances were calculated with the tests indicated in the various figure legends. P values lower than 0.05 were considered statistically significant.

RESULTS AND DISCUSSION

First- and second-generation adenoviral vectors (AdVs) are rendered replication-defective through the removal of only a few viral ORFs (17,38), which prevents exploiting the aforesaid full 36-kb packaging capacity of adenoviral capsids. Moreover, at high vector doses, so-called 'leaky' viral gene expression from the remaining ORFs contributes to cytotoxic effects *in vitro* and immune responses *in vivo* (37). Therefore, we focused on assembling third-generation AdVPs whose vector genomes contain exclusively recombinant DNA (**Supplementary Figure S2**). In particular, recombinant DNA encoding the PE2 fusion protein alone or together with a pegRNA. AdVPs expressing the mCherry reporter were also assembled to monitor transduction efficiencies. Another important aspect to consider concerns the cell tropism of AdVs. Specifically, AdVs with capsids from prototypic adenovirus type-5 enter cells after binding to the Coxsackie B and adenovirus receptor (CAR) (39,40). Yet, scientifically and therapeutically relevant human cell types, e.g., mesenchymal stem cells (hMSCs), muscle progenitor cells (myoblasts); and hematopoietic stem cells, all have a paucity of CAR on their plasmalemmas (41–45). Thus, to test prime editing in both CAR-positive and CAR-negative human cells, recombinant vector constructs were packaged in adenoviral capsids displaying type-50 fibers (**Supplementary Figure S2**), as these fibers have as primary receptor the ubiquitously expressed type I membrane protein CD46 (46).

AdVP.PE2 and AdVP.PE2.pegRNA^{CTT}, encoding PE2 alone and PE2:pegRNA^{CTT} complexes, respectively (**Figure 1A**), were produced to similar high titers and contained structurally intact DNA with evidence neither for rearranged nor truncated species (**Supplementary Figure S3**). PE2:pegRNA^{CTT} complexes insert a CTT triplet at the long non-coding RNA gene *LINC01509*, earlier dubbed HEK293 site 3 (HEK3) (15). Transduction of HeLa cells with AdVP.PE2 and AdVP.PE2.pegRNA^{CTT} yielded widespread and vector dose-dependent PE2 expression (**Figure 1B** and **C**, respectively). Quantification of transduction levels showed that applying AdVPs at a multiplicity of infection (MOI) at or above 3×10^3 vector genome copies per cell (GC $cell^{-1}$) led to transgene expression in virtually all HeLa cells (**Supplementary Figure S4**). Importantly, AdVP.PE2.pegRNA^{CTT} achieved all-in-one delivery of functional prime editing complexes as demonstrated by robust LINC01509 editing in HeLa and HEK293T cells (**Figure 1D** and **E**, respectively). Transfection of the latter cells with plasmids expressing PE2 and pegRNA^{CTT}, yielded prime editing frequencies of 25 ± 1.7% (**Figure 1E**). Transduction of the same easy-to-transfect cells with AdVP.PE2.pegRNA^{CTT} led to significantly higher prime editing frequencies (**Figure 1E**).

In the foundational study, prime editing frequencies were substantially higher in HEK293T cells than in other cell lines tested (15). Furthermore, the less efficient, yet simpler and less mutagenic PE2 system, was not evaluated in cells other than HEK293T cells (15). Physical and chemical transfection methods permit introducing genome editing reagents into cells in a transient fashion. However, reaching maximum delivery efficiencies without triggering substantial cytotoxic effects is challenging, especially

in settings involving non-transformed cells. To start investigating the performance of prime editing in difficult-to-transfect human cells under conditions in which the attendant tools are not limiting, we transduced primary hMSCs (**Figure 2A**) with AdVP.PE2.pegRNA^{CTT} (**Figure 1A**) and AdVP.mCherry (**Supplementary Figure S4A**). A vector dose-dependent build-up of PE2 and mCherry was readily detected through western blotting and direct fluorescence microscopy analyses (**Figure 2B** and **Supplementary Figure S5A**, respectively). Transduction efficiencies, as determined by flow cytometry, varied from a minimum of 92.4 ± 2.2% to a maximum of 99.2 ± 0.8% (**Supplementary Figure S5B**). Importantly, AdVP.PE2.pegRNA^{CTT}-transduced hMSCs contained the intended edits in up to 31% (27.6 ± 3.8%) of target alleles with limited genomic incorporation of complex small insertions and deletions (indels) and pegRNA scaffold sequences (**Figure 2C**–**E**). Indels were found at higher frequencies than scaffold footprints and, together, they reached a combined maximum of 2.05% and 2.80% in hMSCs exposed to the lowest and highest vector concentrations, respectively (**Figure 2C**). Moreover, hMSCs retained their differentiation capacity regardless of the vector dose applied (**Figure 2F**), as determined by quantification of osteogenic and adipogenic differentiation using colorimetric and RT-qPCR assays (**Figure 2G** and **H**, respectively). Taken together, these data indicate that AdVP-based prime editing achieves efficient and precise genetic modification of target alleles in primary hMSCs.

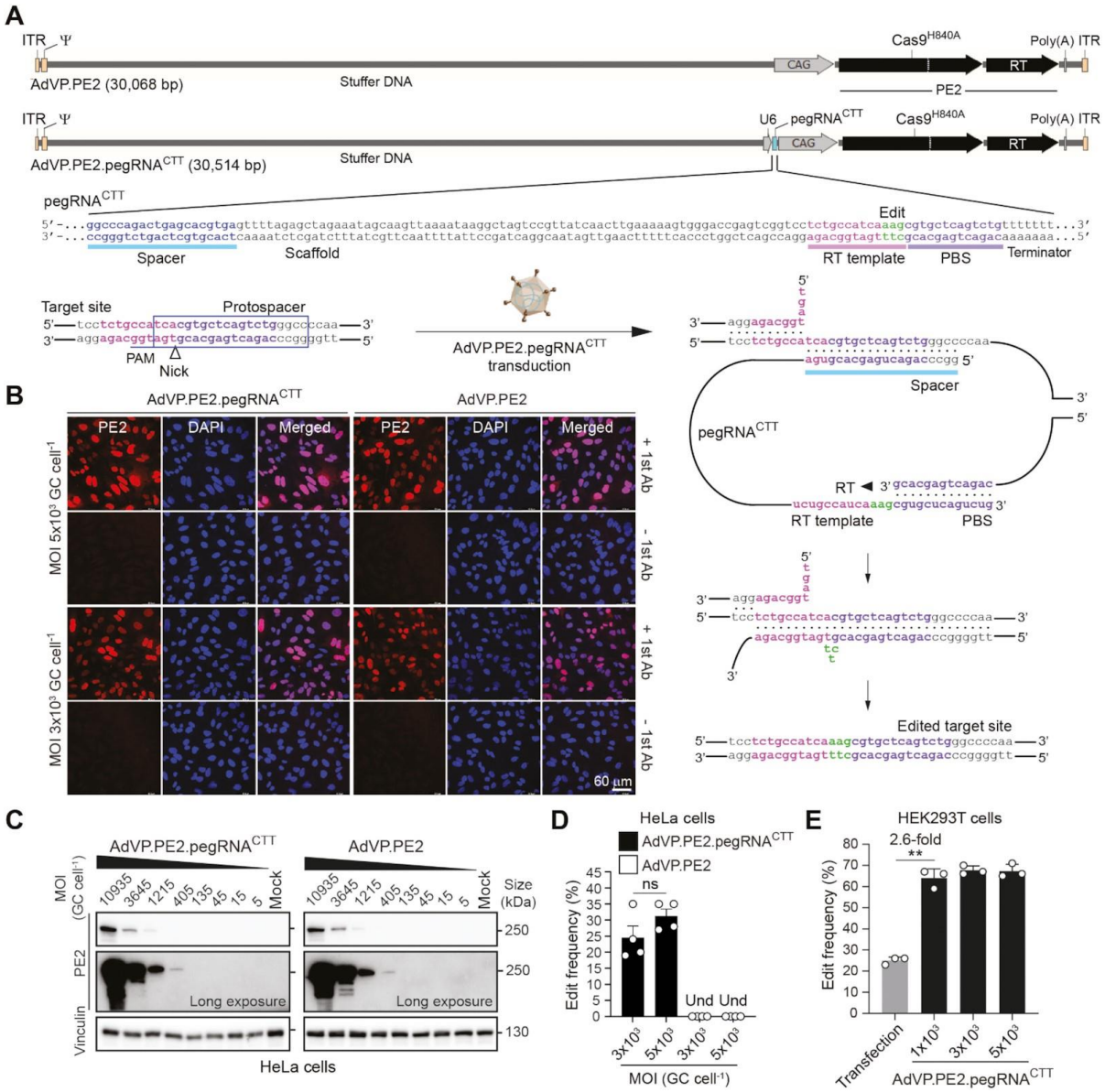


Figure 1. Efficient prime editing in human cells through all-in-one AdVP delivery. (**A**) Schematics of AdVP genomes encoding PE elements. AdVP.PE2 and AdVP.PE2.pegRNA^{CTT} express only PE2 and PE2:pegRNA^{CTT} complexes, respectively. The pegRNA^{CTT} installs a CTT insertion at the long non-coding RNA gene *LINC01509* once coupled to PE2. Hybrid CAG and human U6 promoters drive PE2 and pegRNA^{CTT} synthesis, respectively. The pegRNA^{CTT} coding sequence is depicted with the respective

components highlighted, i.e., sequence-specific spacer, scaffold, primer binding site (PBS) and RT template with embedded edit. A schematics of target site engagement and editing by a PE2:pegRNA^{CTT} complex is equally shown. PAM, protospacer adjacent motif (NGG). (**B** and **C**) Analyses of PE2 expression in transformed cells. HeLa cells were transduced with AdVP.PE2.pegRNA^{CTT} or AdVP.PE2 at the indicated multiplicities-of-infection (MOI). PE2 was detected by immunofluorescence microscopy (panel B) and western blotting (panel C) at 2 days post-transduction. Cas9- and vinculin-specific antibodies served for detecting target (PE2) and loading control proteins, respectively. (**D** and **E**) Detection of prime editing in transformed cells. Cervical carcinoma HeLa cells and human embryonic kidney HEK293T cells were transduced with the indicated AdVPs at different MOI (panels D and E, respectively). At 3 days post-transduction, prime editing activities were assessed by target site genotyping assays. HEK293T cells were also co-transfected with plasmids encoding PE2 and pegRNACTT. GC cell–1, vector genome copies per cell. Und, undetected. Graphs in panels D and E present mean ± s.e.m. (n = 4) and mean ± s.d. (n = 3), respectively. Statistical significance between the indicated datasets was assessed with two-tailed Student's t tests; **0.001<*P*<0.01; *P*> 0.05 was considered non-significant (ns).

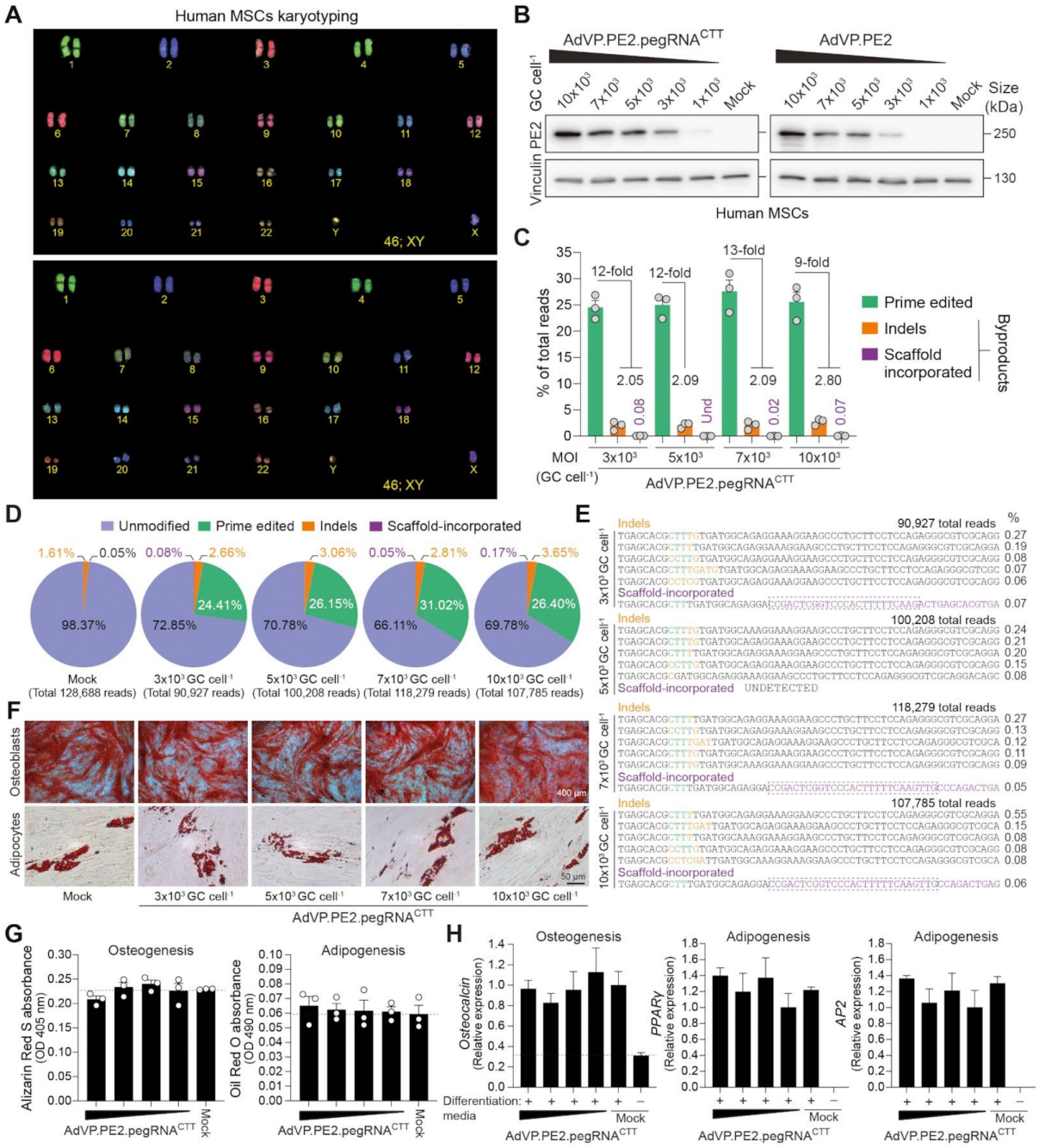


Figure 2. Efficient prime editing in primary human mesenchymal stem cells through all-in-one AdVP delivery. (**A**) Karyotyping of hMSCs. COBRA-FISH analysis was used for confirming the normal diploid status (46;XY) of primary hMSCs. (**B**) Analyses of PE2 expression in primary hMSCs. hMSCs were transduced with AdVP.PE2.pegRNA^{CTT} or AdVP.PE2 at the indicated MOI. At 2 days post-transduction, PE2 expression was assessed by western blotting. Cas9- and vinculin-specific antibodies detected target and loading control proteins, respectively. (**C**) Relationship between prime edited alleles and byproduct variants in AdVP.PE2.pegRNA^{CTT}-transduced hMSCs. hMSCs were transduced with the indicated AdVPs at different MOI. At 3 days post-transduction, prime editing frequencies and unwarranted byproducts (i.e. indels and scaffold-derived insertions), were determined

through CRISPResso2 analysis. Byproducts consist of small insertions and deletions (indels; orange bars) plus insertions derived from reverse transcription into the pegRNA scaffold (violet bars). Graph presents mean ± s.e.m. of three biological replicates; Und, undetected. (**D**) Characterization of prime editing in hMSCs. Pie chart parsing the frequencies of unmodified and modified alleles resulting from a transduction experiment in hMSCs. (**E**) Characterization of prime editing byproducts in hMSCs. Sequences and frequencies of the most frequent alleles bearing indels and pegRNA scaffold-derived insertions from a transduction experiment in hMSCs are presented. (**F**) Differentiation of prime-edited hMSCs. Differentiation capacity of mock- and vector-transduced hMSCs was established after their exposure to defined culture conditions. Osteoblasts and adipocytes were identified by Alizarin Red S and Oil Red O staining, respectively. (**G**) Quantification of osteogenic and adipogenic differentiation through Alizarin Red S and Oil Red O colorimetry, respectively. (**H**) Quantification of osteogenic and adipogenic differentiation via RT-qPCR targeting the indicated lineage-specific marker transcripts. Data are plotted as mean ± s.e.m. of three technical replicates. In all hMSCs transduction experiments MOIs ranged from 3×10^3 through 10×10^3 genome copies per cell (GC cell^{-1}).

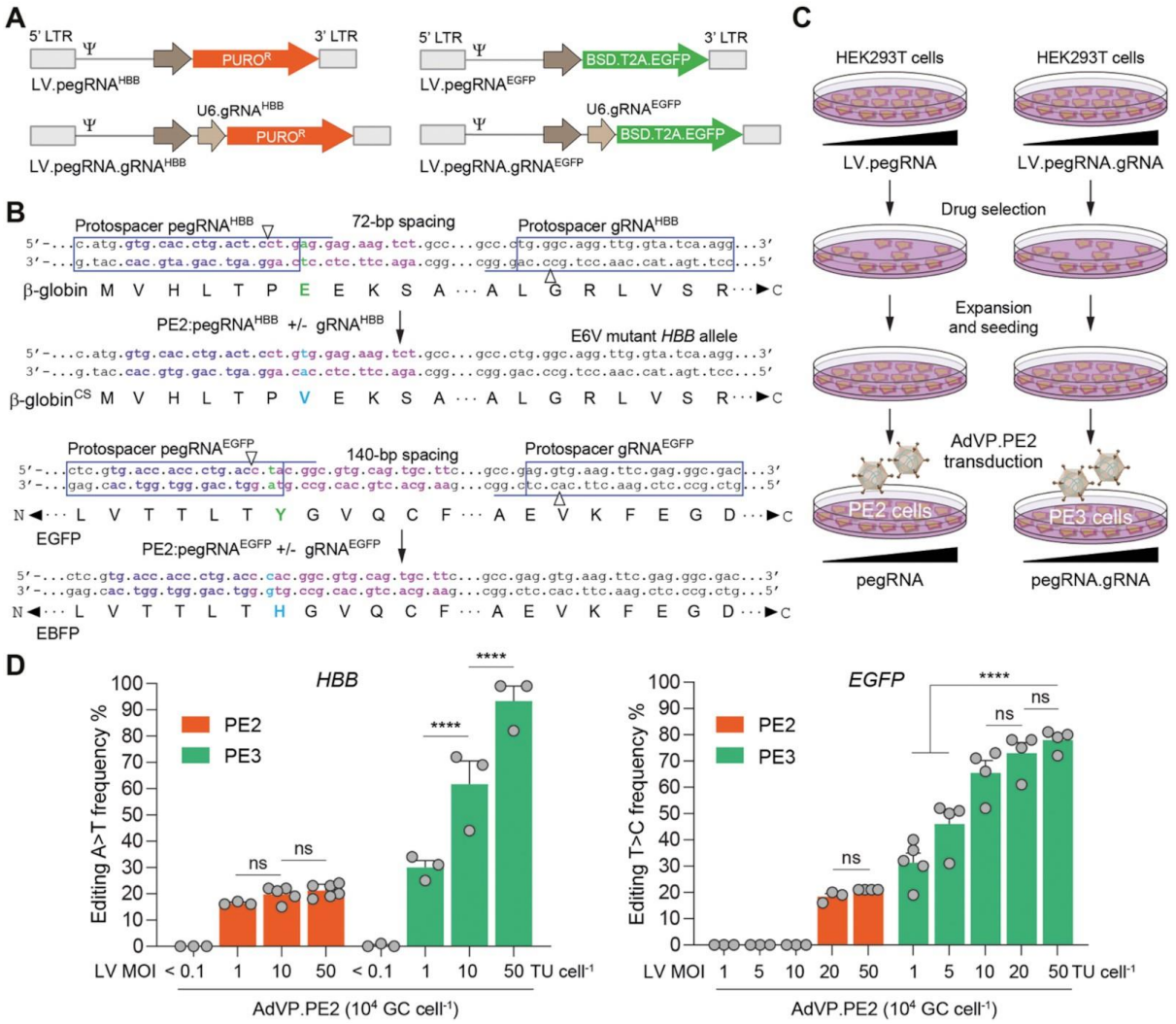


Figure 3. Differential prime editing dose-responses of PE2 versus PE3 systems. (**A**) Lentiviral vectors encoding PE2 and PE3 small RNAs. PE2:pegRNA^{HBB} complexes alone (PE2 system) or together with pairing gRNA^{HBB} (PE3 system) install a A > T transversion at HBB whose β-globin E6V product causes sickle cell (SC) disease. PE2:pegRNA^{EGFP} complexes alone (PE2 system) or together with pairing gRNA^{EGFP} (PE3 system) install a T > C transition, changing the EGFP fluorophore to that of EBFP. *HBB*-targeting lentiviral vectors code for puromycin N-acetyltransferase and *EGFP*-targeting lentiviral vectors code for blasticidin S deaminase and EGFP via a BSD.T2A.EGFP expression unit. LTR and Ψ, long terminal repeat and packaging signal elements, respectively. (**B**) Schematics of *HBB* and *EGFP* target sites before and after prime editing. Protospacers and protospacer adjacent motifs of pegRNAs and gRNAs are boxed and underlined, respectively. Genomic sequences with complementarity to pegRNA primer binding sites and reverse transcriptase templates are marked in violet and magenta, respectively. (**C**) Diagram of the experimental set-up. HEK293T cells stably expressing different amounts of pegRNAs or pegRNA/gRNA pairs were generated by lentiviral vector transduction and puromycin or blasticidin selection. The resulting cell populations were then transduced with AdVP.PE2 at 10^4 GC cell^{-1}. (**D**) Comparing PE2 and PE3 editing efficiencies. PE2- and PE3-induced editing frequencies at endogenous *HBB* and recombinant *EGFP* alleles were determined at three days after AdVP.PE2 transductions. Data shown represent mean ± s.e.m. of at least three independent biological replicates. Significance between the indicated datasets was calculated with two-way ANOVA followed by Sidak's test for multiple comparisons; ****P<0.0001; P>0.05 was considered non-significant (ns).

Interestingly, regardless of target cell type, vector dose-dependent accumulation of PE2 components was not accompanied by a measurable increase in prime editing frequencies (**Figures 1D**, **E** and **2C**). To investigate whether this marked nonlinear dose-response and wide DNA editing plateau is specific to pegRNA^{CTT} or to the PE2 system, we compared PE2 and PE3 systems in HEK293T cells containing individual pegRNAs and pegRNA/gRNA pairs, respectively, delivered through lentiviral vector transductions at various MOI (**Figure 3A**). Cells expressing correspondingly different amounts of these exogenous small RNAs, designed for installing the pathogenic E6V β-globin mutation at *HBB* (15) or for changing the EGFP fluorophore to that of EBFP (**Figure 3B**), were subsequently transduced at high MOI with AdVP.PE2 in order to guarantee uniform PE2 protein availability (**Figure 3C**). A direct correlation between exogenous RNA amounts and prime editing activities was most obvious in cells containing PE3 components with 93% and 78% of *HBB* and *EGFP* sequences on average modified, respectively (**Figure 3D**). These data indicate that, besides being more efficient than the PE2 system on a per dose basis, the PE3 system involving gRNA-directed nicking of the non-edited DNA strand (**Supplementary Figure S1B**), can readily overcome the maximal activity plateau reached by PE2:pegRNA complexes (**Figures 1D**, **E**, **2C** and **3D**). Moreover, populations of cells containing single- to low-copy numbers of pegRNA expression units had negligible frequencies of edited alleles (**Figure 3D**). This finding might have implications for genome-wide screens based on single-copy chromosomal integration of pegRNA libraries in test cell populations. Finally, AdVP transduction experiments in EGFP-expressing HEK293T cells followed by reporter-directed flow cytometry confirmed the emergence of cells with EBFP-positive and EGFP-negative phenotypes expected from the activities of the prime editing reagents herein designed for in cellula fluorophore conversion (**Supplementary Figure S6**).

By capitalizing on the capsid-mediated cell nucleus entry mechanisms of adenoviruses (47,48), recombinant vector forms are proficient in transducing dividing and fully post-mitotic cells (17–20,38). Therefore, we next exploited this evolution fine-tuned feature to investigate whether cell replication influences prime editing activities and outcomes. To start addressing these questions, we used human myoblasts and differentiated post-mitotic myotubes as an experimental model (**Figure 4A**). Irreversible cell cycle withdrawal is a defining feature of the skeletal muscle differentiation program whose regulation involves the MyoD1 family of transcription factors (49). We have confirmed myogenic differentiation of dividing myoblasts into syncytial non-dividing myotubes through combined immunodetection of cell cycling- and late muscle-specific markers (**Supplementary Figure S7A** and **B**). In addition, the presence of actively dividing cells in myoblast cultures as well as their absence in post-mitotic myotube cultures was independently ascertained through a proliferating assay comprising the incorporation of the thymidine analogue EdU in newly synthesized DNA chains (**Supplementary Figure S7B** and **C**).

Similarly to the transduction of HeLa cells and hMSCs (**Supplementary Figures S4** and **S5**, respectively), the transduction of myoblasts with AdVPs was highly efficient (**Supplementary Figure S8**). In addition, exposing skeletal muscle cells to AdVP.PE2.pegRNA^{CTT} at the myoblast and myotube differentiation stages, yielded similar amounts of PE2 protein on a per vector dose basis (**Figure 4B**). Noticeably, the frequencies of precisely edited alleles were highest (up to 49.6%) in myoblasts (**Figure 4C** and **D**), with these muscle progenitors retaining their cell-cycle exit and differentiation capabilities once exposed to poor-mitogen conditions (i.e. no or low serum concentrations) that normally trigger myogenic cell maturation (**Figure 4E** and **Supplementary Figure S9**). In comparison with target alleles containing exclusively the programmed edits, those bearing complex indels and pegRNA scaffold sequences were rare in that the frequencies of these variants combined ranged from 1.3% to 3.3% of the total allelic forms found (**Figure 4C**). In contrast to the results obtained in proliferating myoblasts, in post-mitotic myotubes, alleles exhibiting pegRNA scaffold sequences were not detected (**Figure 4C**) possibly due to the lower prime editing activity in these cells or their non-replicating status (**Supplementary Figure S7**).

The ability to instruct different genomic changes at once contributes to the versatility of the prime editing system. In the myoblast-to-myotube cellular differentiation system, an homozygous single nucleotide polymorphism (SNP) present in the region complementary to the pegRNA^{CTT} RT template was harnessed to investigate the proportions between edited alleles comprising CTT insertions and A > G substitutions (CTT + G) and those containing only one of the two modifications, i.e., CTT or G (**Supplementary Figure S10**). A striking difference between prime editing outcomes in myoblasts and myotubes was the prevalence of the instructed CTT + G composite edits in the former, i.e., over 50% of

the edited reads (**Figure 4F** and **Supplementary Figure S11**). This data suggests that the cellular context has a bearing on the performance of prime editing aiming at simultaneous introduction of different nucleotide changes and extends earlier research showing that the rates of partial edits increase with the distance between the two editing positions and between these positions and the nicking site (50).

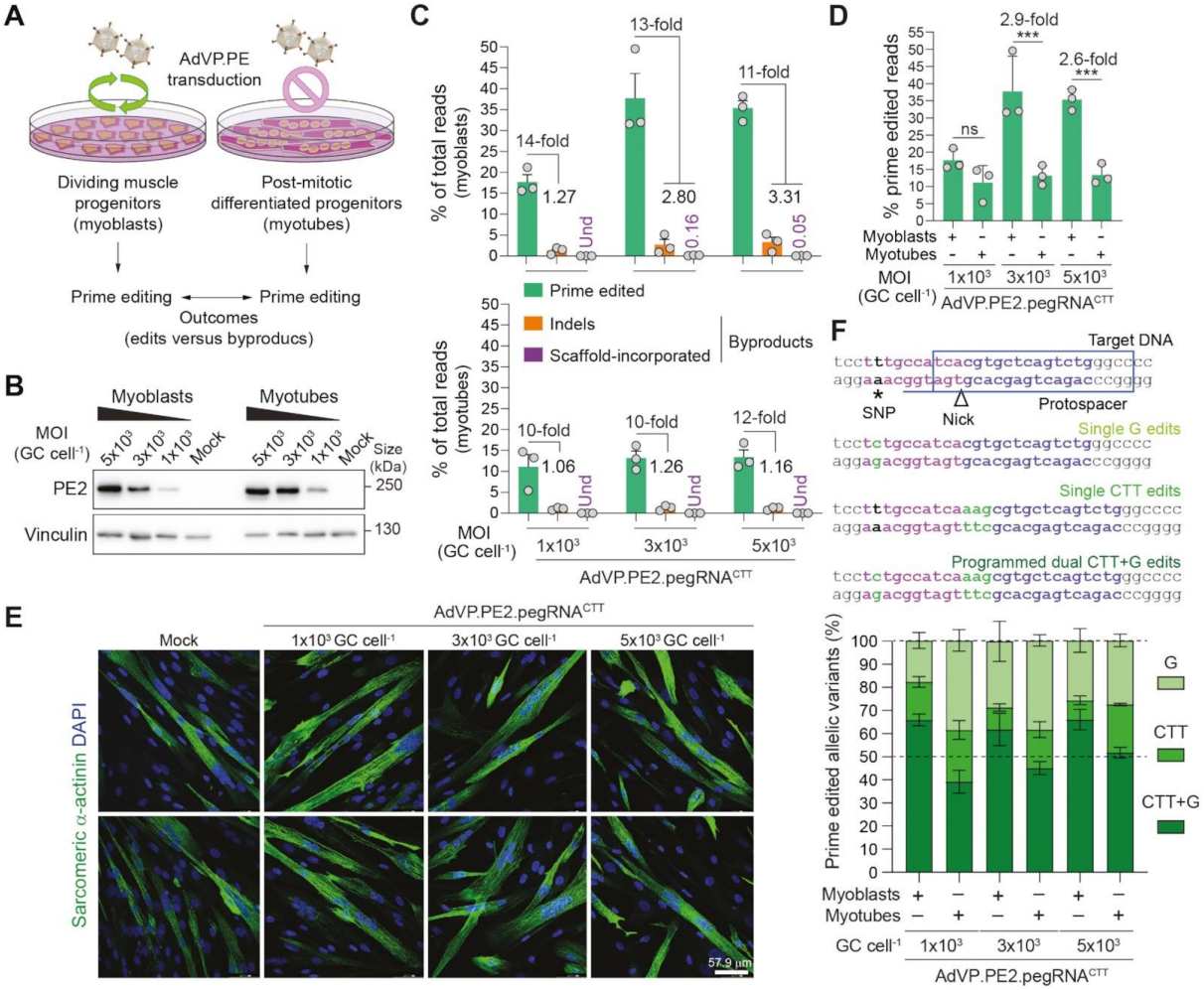


Figure 4. Assessing the impact of cellular replication on prime editing performance. (**A**) Experimental set-up for characterizing prime editing in dividing versus post-mitotic cells. (**B**) Analyses of PE2 expression in muscle cells. Muscle progenitor cells were transduced before and after differentiation with AdVP.PE2.pegRNA^{CTT} at the specified MOI and, 2 days later, PE2 expression levels were assessed by western blotting. Cas9- and vinculin-specific antibodies detected target (PE2) and loading control proteins, respectively. (**C**) Characterization of prime editing in dividing versus post-mitotic cells. The frequencies of prime edited and byproduct alleles in myoblasts and myotubes (top and bottom graphs, respectively) exposed to different MOI of AdVP.PE2.pegRNA^{CTT}, were determined through CRISPResso2 analysis at 3 days post-transduction. Byproducts consist of small deletions and insertions (indels; orange bars) and pegRNA scaffold-derived insertions (violet bars). Graph presents mean ± s.e.m. of three biological replicates; Und, undetected. (**D**) Prime editing frequencies in mitotic versus post-mitotic cells. Aggregated prime editing frequencies indicated in panel C highlighting differences in PE2:pgRNA^{CTT} activity in myoblasts versus myotubes. Bars and error bars correspond to mean and s.d, respectively. Significance between datasets was calculated with two-way ANOVA followed by Sidak's test for multiple comparisons; ***0.0001<*P*<0.001; *P*>0.05 was considered non-significant (ns). (**E**) Differentiation of AdVP.PE2.pegRNA^{CTT}-treated myoblasts. The differentiation capacity of vector-transduced myoblasts was ascertained by immunofluorescence microscopy analysis for the late muscle-specific marker sarcomeric α-actinin after incubating the cells in low-mitogen medium. Nuclei in syncytial myotubes were identified by DAPI staining. Mock-transduced myoblasts served as controls. Two representative micrographs for each experimental condition are shown. (**F**) Parsing of prime-edited allele variants resulting from a composite prime editing design. Myoblasts and RPE-Fucci cells are homozygous for a SNP in the region complementary to the pegRNA^{CTT} RT template permitting assessing composite (CTT + G) versus single (CTT or G) edits instructed by PE2:pegRNA^{CTT} complexes. Genomic sequences before and after the delivery of PE2:pegRNA^{CTT} complexes into cells containing a SNP in the region complementary to the RT template (magenta nucleotides), are depicted (top panel). Prime editing outcomes instructed by pegRNA^{CTT} correspond to alleles containing A > G substitutions, CTT insertions or both modifications. Replicating myoblasts and post-mitotic myotubes were transduced with the all-in-one vector AdVP.PE2.pegRNACTT. Discrimination and quantification of the different target alleles was performed via next-generation deep sequencing analysis on genomic DNA isolated at 3 days post-transduction (bottom panel). Data are plotted as mean ± s.e.m. of three biological replicates.

To further exploit AdVPs as probes for investigating the relationship between cellular replication and prime editing, we next used retinal pigment epithelial cells (RPE-1) endowed with the Fucci cell-cycle sensors (RPE-Fucci) (**Figure 5A**) (26,28,29). In contrast to most transformed cells, non-transformed RPE-1 cells have functional G1 and G2 checkpoints (28,29). Most importantly, the Fucci system allowed for flow cytometric quantification of cell fractions in G1, G2/M and early S phases in cultures treated with serum concentrations spanning from mitogen starvation to regular conditions. During the cumulative tracing period, a clear serum concentration-dependent increase in the proportion of cells in S and G2/M was observed (**Figure 5B** and **Supplementary Figure S12A**). This observation was independently confirmed by assessing throughout the same cumulative period, the DNA content profiles of RPE-Fucci cells exposed to different serum concentrations (**Supplementary Figure S13**).

AdVP.PE2.pegRNA^{CTT} transduction of RPE-Fucci cells treated with these various serum concentrations contained similar amounts of PE2 (**Figure 5C**). Significantly, prime editing efficiencies were superior in RPE-Fucci populations with the highest frequencies of actively cycling cells (**Figure 5D**). Interestingly, within the fraction of edited alleles, the frequencies of programmed CTT + G alleles were higher than those corresponding to partially edited CTT or G alleles, especially so in the most actively replicating cell populations (**Figure 5E** and **Supplementary Figure S12B**). The direct correlation between RPE-Fucci cell-cycling activities and composite CTT + G allele frequencies, is in line with the data obtained after transduction of myoblasts and post-mitotic myotubes with AdVPs as a higher purity of the programmed CTT + G edits was observed in the proliferating myoblast populations (**Figure 4F**). Equally consistent with the experiments in muscle cells, complex indels and genomic insertions mapping to pegRNA sequences were rarer in cultures enriched in G1-arrested cells (**Figure 5F** and **Supplementary Figure S12B**). In general, the latter prime editing byproducts were the least frequent and were prevalently found in cultures containing high frequencies of mitotically active cells (**Figures 2C**–**E**, **4C**, **5F** and **Supplementary Figures S11** and **S12B**).

Taken together, our data demonstrate that the replication status of target cells influences prime editing activities and outcomes. Edited strand-templated DNA synthesis (**Supplementary Figure S14**) and/or engagement of cellular factors involved in replication-dependent DNA repair processes might contribute to the herein identified correlation between replication and prime editing activities. Prime editing and non-LTR element retrotransposition share striking similarities, which include genomic DNA nicking followed by target site-primed reverse transcription (51). This fact coupled to the copious abundance of non-LTR retrotransposons in mammalian genomes makes it equally enticing to speculate the participation of conserved retrotransposition-associated processes during prime editing.

Recently, dual AAV strategies involving co-transducing target cells with two split AAV vectors, each encoding N- or C-terminal truncated prime editors, were applied for in situ reconstruction of full-length proteins through vector genome recombination and RNA trans-splicing (52–54) or intein-mediated protein trans-splicing (55,56). These studies yielded important proof-of-concepts for in vivo disease modelling and mutation correction by prime editing. In particular, a dual AAV protein trans-splicing PE3 system corrected 0.6%, 2.3% and 3.1% of defective alleles in livers of alpha-1 antitrypsin deficiency mice at 2, 6 and 10 weeks post-injection, respectively (57). Another study reports that, in the absence of positive selection for gene-corrected cells, dual AAV RNA trans-splicing PE2 and PE3 systems did not lead to detectable correction of defective alleles in livers of hereditary tyrosinemia type 1 mice (58). Notably, however, using the same AAV platform, gene correction levels ranging from 4.1% to 7.4% were measured in RPE65-associated Leber congenital amaurosis mice after sub-retinal injections (58). In another recent study, a dual AAV protein trans-splicing system installed a G > T transversion within Dnmt1 at a frequency of 1.71 ± 1.35% 6 weeks after sub-retinal injections in mice (59). Notwithstanding these important proof-of-concepts, dual AAV designs are complex and relatively inefficient as they operate through only partially controllable intermolecular recombination events for assembling the proper full-length product from the total pool of truncated proteins expressed in co-transduced cells. There is, therefore, a pressing need for additional platforms capable of delivering into different human cell types and animal models emerging RNA-guided gene targeting systems based on large macromolecular complexes, as these systems are starting to offer the prospect for unprecedented genome editing precision in fundamental research, disease modelling and treatment (16).

In this work, besides identifying a direct correlation between prime editing and cell cycling activities, we demonstrate that the AdVP platform is a suitable option for all-in-one delivery of full-length prime editor proteins and cognate RNA molecules into target cells independently of their transformation and replication statuses. Hence this platform is expected to become a valuable addition for investigations on the potential and limitations of prime editing principles and reagents in a broad range of mammalian cells *in vitro* and *in vivo*.

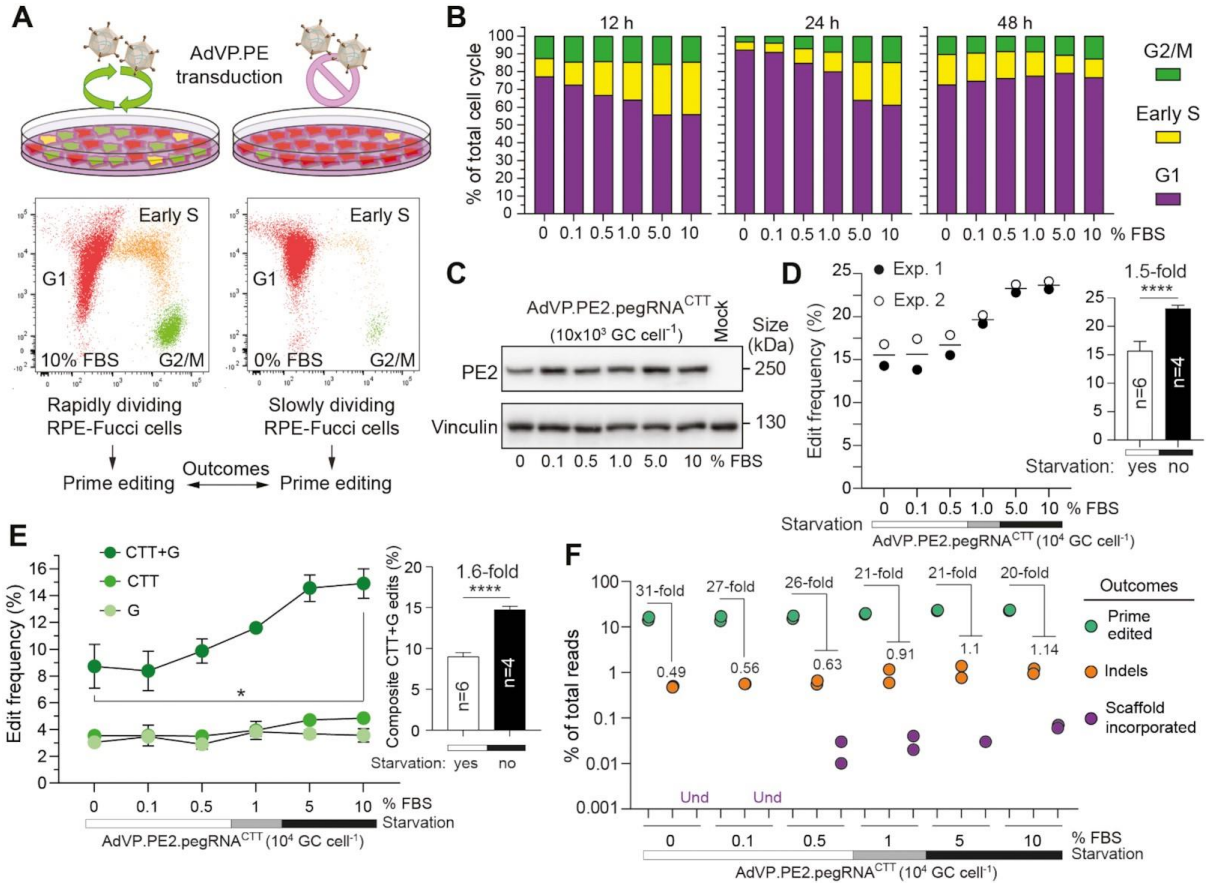


Figure 5. Characterization of prime editing in rapidly versus slowly dividing cell populations. (**A**) Experimental set-up for characterizing prime editing in rapidly versus slowly dividing cell populations. RPE-Fucci cell cultures containing varying proportions of cycling and non-cycling cells were established by applying a serum gradient. Fucci reporters in these cells trace cell fractions in G1, early S or G2/M phases, permitting the monitoring of cell division-to-prime editing rates upon AdVP.PE2.pegRNA^{CTT} transduction. (**B**) Tracking cell cycling and prime editing activities in RPE-Fucci cell cultures. Cell cycle analysis was done at the indicated timepoints on RPE-1 cells exposed to various FBS concentrations by flow cytometry. Parallel cultures of RPE-Fucci cells were transduced with AdVP.PE2.pegRNA^{CTT} at 12 hours after FBS treatments initiation. (**C**) Analysis of PE2 expression in RPE-1 cells. PE2 expression levels were determined by western blotting using Cas9- and vinculin-specific antibodies for detecting PE2 and loading control proteins, respectively. (**D**) Quantification of prime editing outcomes in RPE-1 cells. Frequencies of prime editing and prime editing collateral events were measured by CRISPresso2 analysis at 60 hours post-transduction. Datapoints derive from two independent experiments carried out throughout different days. Bars represent mean ± s.d. of prime editing activities in RPE-Fucci cells exposed to serum starvation (0, 0.1, and 0.5% FBS) or regular culture conditions (5 and 10% FBS). Unpaired two-tailed Student's t test ****P<0.0001. (**E**) Parsing of prime-edited allele variants resulting from the composite prime editing design. Statistical significance was assesed by Student's t tests; *0.01<P<0.05; ****P< 0.0001. (**F**) Relationship between prime edited alleles and byproduct variants in AdVP-transduced RPE-Fucci cells. Und, undetected.

DATA AVAILABILITY

All data gathered for and analysed in this study are included in the article and supplementary files. Additional raw datasets that support the findings of this work are available upon request. The flow cytometry datasets were deposited at FlowRepository under accession codes FR-FCM-Z3RU, FR-FCM-Z3RV, FR-FCM-Z4GP, FR-FCM-Z4GQ and FR-FCM-Z4GX. The generated deep sequencing library reads were deposited at the NCBI Sequence Read Archive (SRA) database under BioProject ID PRJNA729964.

SUPPLEMENTARY DATA

Supplementary Data are available online at doi: 10.1093/nar/gkab938.

ACKNOWLEDGMENTS

The authors thank Martijn Rabelink and Steve Cramer for performing $p24^{gag}$ ELISA titrations of lentiviral vector particles and assistance with the research logistics, respectively. The authors also thank Karoly Szuhai for the FISH karyotyping, Ignazio Maggio for input on the amplicon deep sequencing procedures, Maarten van Dinther and Midory Thorikay for assistance with the hMSC differentiation assays, Chuannan Fan for his help with the Ki-67 and dystrophin western blot assays (Department of Cell and Chemical Biology, LUMC), and Vincent Mouly from the Center for Research in Myology (Sorbonne University, Paris, France) for providing the human myoblasts. Finally, the authors are grateful to all laboratory members for their support.

FUNDING

Prinses Beatrix Spierfonds (W.OR16-13, W.OR21-01); Dutch Duchenne Parent Project (17.012); China Scholarship Council–Leiden University Joint Scholarship programme (to Q.W.). This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 765269 (to F.T.). Funding for open access charge: Prinses Beatrix Spierfonds.

Conflict of interest statement. None declared.

REFERENCES

1. Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. (2012) *Proc. Natl. Acad. Sci. USA*, **109**, E2579-2586.
2. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* **337**, 816-821.
3. Cradick, T.J., Fine, E.J., Antico, C.J. and Bao, G. (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*., **41**, 9584–9592.
4. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol*., **31**, 822–826.
5. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol*., **31**, 827–832.
6. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Bioatechnol*., **31**, 839–843.
7. Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J. and Bao, G. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res*., **42**, 7473–7485.
8. Wang, Q., Liu, J., Janssen, J.M., Le Bouteiller, M., Frock, R.L. and Gonçalves, M.A.F.V. (2021) Precise and broad scope genome editing based on high-specificity Cas9 nickases. *Nucleic Acids Res*., **49**, 1173-1198.
9. Turchiano, G., Andrieux, G., Klermund, J., Blattner, G., Pennucci, V., El Gaz, M., Monaco, G., Poddar, S., Mussolino, C., Cornu, T.I. et al. (2021) Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell*, **28**, 1136-1147.
10. Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E. and Alt, F.W. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol*., **33**, 179-186.
11. Kosicki, M., Tomberg, K. and Bradley, A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**, 765-771.
12. Chen, X., Janssen, J.M., Liu, J., Maggio, I., 't Jong, A.E.J., Mikkers, H.M.M. and Gonçalves, M.A.F.V. (2017) In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun*., **8**, 657.
13. Nakajima, K., Zhou, Y., Tomita, A., Hirade, Y., Gurumurthy, C.B. and Nakada, S. (2018) Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res*., **28**, 223–230.

14. Hyodo, T., Rahman, M.L., Karnan, S., Ito, T., Toyoda, A., Ota, A., Wahiduzzaman, M., Tsuzuki, S., Okada, Y., Hosokawa, Y. et al. (2020) Tandem paired nicking promotes precise genome editing with scarce interference by p53. *Cell Rep*., **30**, 1195–1207.
15. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. et al. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
16. Anzalone, A.V., Koblan, L.W. and Liu, D.R. (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824-844.
17. Chen, X. and Gonçalves, M.A.F.V. (2016) Engineered viruses as genome editing devices. *Mol. Ther*., **24**, 447-457.
18. Jager, L., Hausl, M.A., Rauschhuber, C., Wolf, N.M., Kay, M.A. and Ehrhardt, A. (2009) A rapid protocol for construction and production of high-capacity adenoviral vectors. *Nat. Protoc.*, **4**, 547-564.
19. Tasca, F., Wang, Q. and Goncalves, M (2020). Adenoviral Vectors Meet Gene Editing: A Rising Partnership for the Genomic Engineering of Human Stem Cells and Their Progeny. *Cells*, **9**, 953.
20. Ricobaraza, A., Gonzalez-Aparicio, M., Mora-Jimenez, L., Lumbreras, S. and Hernandez-Alcoceba, R. (2020) High-capacity adenoviral vectors: expanding the scope of gene therapy. *Int. J. Mol. Sci.*, **21**, 3643.
21. Holkers, M., Maggio, I., Henriques, S.F., Janssen, J.M., Cathomen, T. and Gonçalves, M.A.F.V. (2014) Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat. Methods*, **11**, 1051-1057.
22. Brescia, M., Janssen, J.M., Liu, J. & Gonçalves, M. (2020) High-capacity adenoviral vectors permit robust and versatile testing of DMD gene repair tools and strategies in human cells. *Cells*, **9**, 869.
23. Chen, X., Rinsma, M., Janssen, J.M., Liu, J., Maggio, I. and Gonçalves, M.A.F.V. (2016) Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res*., **44**, 6482-6492.
24. Gonçalves, M.A.F.V., Swildens, J., Holkers, M., Narain, A., van Nierop, G.P., van de Watering, M.J., Knaän-Shanzer, S. and de Vries, AA. (2008) Gonçalves, M.A. et al. Genetic complementation of human muscle cells via directed stem cell fusion. *Mol. Ther.*, **16**, 741-748.
25. Gonçalves, M.A.F.V., Janssen, J.M., Nguyen, Q.G., Athanasopoulos, T., Hauschka, S.D., Dickson, G. and de Vries A.A. (2011) Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. *Mol. Ther*., **19**, 1331-1341.
26. Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H. et al. (2008) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell*, **132**, 487-498.
27. Mamchaoui, K., Trollet, C., Bigot, A., Negroni, E., Chaouch, S., Wolff, A., Kandalla, P.K., Marie, S., Di Santo, J., St Guily, J.L. (2011) Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. *Skelet Muscle*, **1**, 34.
28. Krenning, L., Feringa, F.M., Shaltiel, I.A., van den Berg, J. and Medema, R.H. (2014) Transient activation of p53 in G2 phase is sufficient to induce senescence. *Mol. Cell*, **55**, 59-72.
29. Shaltiel, I.A., Aprelia, M., Saurin, A.T., Chowdhury, D., Kops, G.J., Voest, E.E. and Medema, R.H. (2014) Distinct phosphatases antagonize the p53 response in different phases of the cell cycle. *Proc. Natl. Acad. Sci. USA*, **111**, 7313-7318.
30. Janssen, J.M., Liu, J., Skokan, J., Goncalves, M.A. and de Vries, A.A. (2013) Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. *J. Gene. Med.*, **15**, 1-11.
31. Havenga, M.J., Holterman, L., Melis, I., Smits, S., Kaspers, J., Heemskerk, E., van der Vlugt, R., Koldijk, M., Schouten, G.J., Hateboer, G. et al. (2008) Havenga, M.J. et al. Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: high yield and preserved bioactivity. *Biotechnol. Bioeng.*, **100**, 273-283.
32. Holkers, M., Cathomen, T. and Goncalves, M.A.F.V. (2014) Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells. *Methods*, **69**, 179-187.
33. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E. et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.*, **37**, 224-226.

34. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*., **17**, 10-12.
35. Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R. and Moriarity, B.S. (2018) EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.*, **1**, 239-250.
36. Dautzenberg, I.J.C., Rabelink, M. and Hoeben, R.C. (2021) The stability of envelope-pseudotyped lentiviral vectors. *Gene Ther.*, **28**, 89-104.
37. Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L. and Trono, D. (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.*, **72**, 9873-9880.
38. Gonçalves, M.A.F.V. and de Vries, A.A. (2006) Adenovirus: from foe to friend. *Rev. Med. Virol*., **16**, 167–186.
39. Bergelson, J.M., Cunningham, J.A., Droguett, G., Kurt-Jones, E.A., Krithivas, A., Hong, J.S., Horwitz, M.S., Crowell, R.L. and Finberg, R.W. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science*, **275**, 1320–1323.
40. Tomko, R.P., Xu, R. and Philipson, L. (1997) HCAR and MCAR: The human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. *Proc. Natl. Acad. Sci. USA*, **94**, 3352–3356.
41. Shayakhmetov, D.M., Papayannopoulou, T., Stamatoyannopoulos, G. and Lieber, A. (2000) Efficient gene transfer into human CD34(+) cells by a retargeted adenovirus vector. *J. Virol.*, **74**, 2567–2583.
42. Knaän-Shanzer, S., van der Velde, I., Havenga, M.J., Lemckert, A.A., de Vries, A.A. 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.
43. Knaän-Shanzer, S., van de Watering, M.J., van der Velde, I., Gonçalves, M.A., Valerio, D. and de Vries, A.A. (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.
44. Gonçalves, M.A., de Vries, A.A., Holkers, M., van de Watering, M.J., van der Velde, I., van Nierop, G.P., Valerio, D. and Knaän-Shanzer, S. (2006) Human mesenchymal stem cells ectopically expressing full-length dystrophin can complement Duchenne muscular dystrophy myotubes by cell fusion. *Hum. Mol. Genet.*, **15**, 213–221.
45. Gonçalves, M.A., Holkers, M., Cudré-Mauroux, C., van Nierop, G.P., Knaän-Shanzer, S., van Der Velde, I., Valerio, D. and de Vries, A.A. (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.
46. Gaggar, A., Shayakhmetov, D.M. and Lieber, A. (2003) CD46 is a cellular receptor for group B adenoviruses. *Nat. Med.*, **9**, 1408–1412.
47. Greber, U.F. and Flatt, J.W. (2019) Adenovirus entry: from infection to immunity. *Annu. Rev. Virol*., **6**, 177-197.
48. Greber, U.F. and Gomez-Gonzalez, A. (2021) Adenovirus - a blueprint for gene delivery. *Curr. Opin. Virol.*, **48**, 49-56.
49. Kitzmann, M. and Fernandez, A. (2001) Crosstalk between cell cycle regulators and the myogenic factor MyoD in skeletal myoblasts. *Cell. Mol. Life Sci*., **58**, 571-579.
50. Kim, H.K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S. and Kim, H.H. (2021) Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol*., **39**, 198-206.
51. Beck, C.R., Garcia-Perez, J.L., Badge, R.M. and Moran, J.V. (2011) LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.*, **12**, 187-215.
52. Nakai, H., Storm, T.A. and Kay, M.A. (2000) Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. *Nat. Biotechnol*., **18**, 527-532.
53. Sun, L., Li, J. and Xiao, X. (2000) Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. *Nat. Med*., **6**, 599-602.
54. Yan, Z., Zhang, Y., Duan, D. and Engelhardt, J.F. (2000) Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proc. Natl. Acad. Sci. USA*, **97**, 6716-6721.
55. Li, J., Sun, W., Wang, B., Xiao, X. and Liu, X.Q. (2008) Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy. *Hum. Gene Ther*., **19**, 958-964.

56. Truong, D.J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W. and Ortiz, O. (2015) Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res*., **43**, 6450-6458.
57. Liu, P., Liang, S.Q., Zheng, C., Mintzer, E., Zhao, Y.G., Ponnienselvan, K., Mir, A., Sontheimer, E.J., Gao G., Flotte, T.R. et al. (2021) Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun*., **9**, 2121.
58. Jang, H., Jo, D.H., Cho, C.S., Shin, J.H., Seo, J.H., Yu, G., Gopalappa, R., Kim, D., Cho, S.R., Kim, J.H. et al. (2021) Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. *Nat. Biomed. Eng*. doi: 10.1038/s41551-021-00788-9. Online ahead of print.
59. Zhi, S., Chen, Y., Wu, G., Wen, J., Wu, J., Liu, Q., Li, Y., Kang, R., Hu, S., Wang, J. et al. (2021) Dual-AAV delivering split prime editor system for in vivo genome editing. *Mol. Ther*. **5**, 283-294.

Chapter 6

Selection-free precise dystrophin repair using high-capacity adenovector delivery of advanced prime editing systems rescues dystrophin synthesis in DMD muscle cells

Qian Wang[1], Sabrina Capelletti[1], Jin Liu[1], Josephine M. Janssen[1] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands

Nucleic Acids Research, 2024, Vol.52, No.5, ***2740–2757***
doi: 10.1093/nar/gkae057

ABSTRACT

Prime editors have high potential for disease modelling and regenerative medicine efforts including those directed at the muscle-wasting disorder Duchenne muscular dystrophy (DMD). However, the large size and multicomponent nature of prime editing systems pose substantial production and delivery issues. Here, we report that packaging optimized full-length prime editing constructs in adenovector particles (AdVPs) permits installing precise *DMD* edits in human myogenic cells, namely, myoblasts and mesenchymal stem cells (up to 80% and 64%, respectively). AdVP transductions identified optimized prime-editing reagents capable of restoring *DMD* reading frames of ~14% of patient genotypes and restore dystrophin synthesis and dystrophin-β-dystroglycan linkages in unselected DMD muscle cell populations. AdVPs were equally suitable for correcting DMD iPSC-derived cardiomyocytes and delivering dual prime editors tailored for *DMD* repair through targeted exon 51 deletion. Moreover, by exploiting the cell cycle-independent AdVP transduction process, we report that 2- and 3-component prime-editing modalities are both most active in cycling than in post-mitotic cells. Finally, we establish that combining AdVP transduction with seamless prime editing allows for stacking chromosomal edits through successive delivery rounds. In conclusion, AdVPs permit versatile investigation of advanced prime editing systems independently of their size and component numbers, which should facilitate their screening and application.

INTRODUCTION

Programmable nucleases consisting of sequence-tailored guide RNAs (gRNAs) and Cas9 endonucleases are powerful tools for genome editing. Yet, the prevalent repair of double-strand DNA breaks (DSB) by error-prone end joining processes confers an intrinsically high mutagenic character to nuclease-based genome editing. In contrast, prime editing permits installing any single base-pair change and precise small insertions or deletions (indels) at specific genomic sequences without DSB formation (1). Typically, prime editing complexes comprise an engineered reverse transcriptase (RT) fused to a nicking Cas9 variant (prime editor) and a 3' end-extended gRNA, named prime editing guide RNA (pegRNA). The pegRNA instructs both target site selection and an edit-of-interest via its spacer and RT template moieties, respectively. Upon target site nicking, annealing of the released single-stranded DNA to the primer binding site (PBS) of the pegRNA primes RT-mediated copying of the RNA template into a complementary DNA which, upon genomic site hybridization, flap excision, and DNA repair or replication, leads to targeted chromosomal edition (1). Prime editing has two main modalities, i.e., PE2 and PE3 (1). The former 2-component system depends solely on a prime editor protein (e.g., PE2) and a pegRNA whilst the latter 3-component system requires a supplementary regular gRNA. In PE3, gRNA-directed nicking of the non-edited DNA strand fosters its replacement by the edited strand which typically results in higher frequencies of homoduplex DNA edits despite a concomitant increase in indel by-products (1). More recently, multiplexing prime editing based on the delivery of prime editors together with dual pegRNAs is contributing to further expand the scope of DSB-independent genome editing procedures. Indeed, in this case, pairs of prime editing complexes act in concert to install genomic insertions, deletions and/or substitutions whose sizes are substantially larger than those enabled through PE2 and PE3 strategies (2-7).

Owing to their vast potential and versatility, prime editing systems are developing at a fast pace and include improved prime editor proteins and pegRNAs, e.g., PEmax (8) and engineered pegRNA (epegRNA) architectures (9,10). The PEmax construct incorporates specific mutations and codon-optimizations in its Cas9 nickase and RT portions, respectively, that contribute to enhanced prime editing activity (8). The epegRNAs have extended 3' ends in the form of structured RNA pseudoknots (e.g., tevopreQ1) that protect them from exonucleolytic degradation (9,10). Notwithstanding these important developments, the large size of prime editing components creates substantial production and delivery bottlenecks that hinder their most efficacious testing and application. Approaches aiming at ameliorating the delivery bottleneck include splitting prime editor constructs in subunits that, upon cell entry, assemble in situ tethered or untethered Cas9 nickase and RT portions (11-20). In addition, other ancillary approaches permit enriching for prime-edited cell fractions via; (i) using surrogate reporter- or drug-based systems for isolating cells co-edited at target and selectable-marker genes (21-23), or (ii) interfering with edited DNA strand removal by co-delivering dominant-negative factors of the cellular DNA mismatch repair pathway (8,10). Although applicable to specific settings, the multicomponent character of these prime-editing systems makes their design complex and their wider application challenging.

High-capacity adenoviral vector particles (AdVPs) form a powerful gene delivery system owing to their extensive packaging capacity (i.e., up to 36-kb), lack of cytotoxic viral genes, high genetic stability, and efficient transduction of dividing and post-mitotic cells (24-26). Indeed, in earlier work, our laboratory has shown that AdVPs allow for combined delivery of regular PE2 components into human cells regardless of their transformation and replication statuses (27). In this study, we investigate the potential of AdVPs for transferring optimized PE2 and PE3 components or optimized prime editing multiplexes for gene repair purposes, namely, for correcting defective *DMD* alleles underlying Duchenne muscular dystrophy (DMD). DMD (MIM #310200) is a X-linked progressive muscle-wasting disorder (incidence: ~1:5,500 boys) caused by loss-of-function mutations in the large *DMD* gene (~2.2 Mb) whose product, dystrophin (427 kDa), plays key structural and physiological roles in striated muscle (28). Interestingly, most DMD-causing mutations consist of intragenic deletions spanning single or multiple exons that disrupt the reading frame. Of notice, in-frame *DMD* deletions yield internally truncated dystrophins whose partial functionality underlies Becker muscular dystrophy (BMD) (MIM #300376), a less acute form of muscular dystrophy. Hence, restoration of the *DMD* reading frame in muscle cells is expected to result in Becker-like dystrophins with therapeutic potential (28).

We report that combining AdVP with improved prime editing systems achieves robust *DMD* gene repair and knockout in muscle progenitor cells (myoblasts) derived from DMD patients and healthy donors, respectively. Indeed, AdVP-assisted restoration of the *DMD* reading frame in human myoblasts with DMD.Δ48-50 and DMD. Δ45-50 genotypes and in induced pluripotent stem cells (iPSCs)-derived cardimyocytes with the latter genotype, readily led to the detection of Becker-like mRNA transcripts and corresponding dystrophin proteins in unselected cell populations. Importantly, proximity ligation assays revealed that the resulting Becker-like dystrophin proteins were capable of connecting to β-dystroglycan, a key member of the dystrophin-associated glycoprotein complex (DGC) present along at the sarcolemma of normal muscle cells. Complementary *DMD* gene repair experiments demonstrated the feasibility and potential of AdVP-based multiplexing prime editing involving all-in-one transfer of optimized full-length prime editor and dual pegRNA components. Moreover, AdVP transduction experiments in cycling myoblasts versus post-mitotic syncytial myotubes established that both PE2 and PE3 systems are most active in dividing cells. Finally, we explored the straightforward AdVP delivery process and the non-mutagenic character of prime editing, to build-up chromosomal edits in target cell populations through consecutive transduction cycles.

MATERIAL AND METHODS

Cells

The human iPSC line CENSOi001-B (herein named DMD iPSCs) used in this study and elsewhere (29) were generate by using an mRNA-based reprogramming protocol on fibroblasts isolated from a DMD patient with a *DMD* deletion spanning exons 45-50. These cells were purchased from the European Bank for induced pluripotent Stem Cells (EBiSC). The DMD iPSCs were maintained in mTeSR medium (STEMCELL Technologies; Cat. No.: 85850) supplemented with 25 U ml^{-1} penicillin/streptomycin (Thermo Fisher Scientific; Cat. No.: 15140122) and cultured in plates coated with Matrigel (Corning Matrigel hESC-Qualified Matrix; Corning; Cat. No.: 354277) according to the manufacturer's guidelines. When 70-80% confluence was reached, the iPSCs were washed with phosphate-buffered saline (PBS) solution (pH 7.4) and then incubated with 0.5 mM ethylenediaminetetraacetic acid (EDTA; Invitrogen Cat. No.: 15575020) in PBS at 37°C for 5 min. After the removal of the EDTA solution, the cells were seeded in mTeSR medium supplemented with a 1:200 dilution of RevitaCell (ThermoFisher Scientific; Cat. No.: A2644501).

HEK293T cells (American Type Culture Collection) were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific; Cat. No.: 41966-029) containing 10% fetal bovine serum (FBS; Biowest; Cat. No.: S1860-500). The AdVP packaging cell line PER.tTA.Cre43 (30), was kept in high-glucose DMEM supplemented with 10% FBS, 10 mM $MgCl_2$ and 0.4 µg ml^{-1} puromycin (Thermo Fisher Scientific; Cat. No.: A11138-03). The characterization of human myoblasts derived from a healthy donor and DMD patients harboring *DMD* intragenic deletions D48–50 or D45–50, herein named, DMD.D45–50 myoblasts (KM1315), DMD.D48–50 myoblasts (AB1098) and DMD.D48–50 myoblasts (6594), have been previously detailed (31,32). These muscle progenitor cells were maintained in Ham's F-10 Nutrient Mixture (Thermo Fisher Scientific; Cat. No.: 41550-021) containing 20% heat-inactivated FBS (Thermo Fisher Scientific; Cat. No.: 10500064), 10 ng ml^{-1} recombinant

human FGF-basic (154 a.a.) (Peprotech; Cat. No.: 100-18B-500-UG), 1 μM Dexamethasone (Sigma-Aldrich; Cat. No.: D2915-100MG) and 100 U ml^{-1} penicillin/streptomycin (Thermo Fisher Scientific; Cat. No.: 15140122). The characterization and culturing of the human mesenchymal stem cells (hMSCs) was detailed elsewhere (33). In brief, these cells were kept in Minimum Essential Medium α (MEM-α) (Thermo Fisher Scientific; Cat. No.: 22561-021) supplemented with 10% FBS, 5 ng ml^{-1} recombinant human FGF-basic (154 a.a.), 100 U ml^{-1} penicillin/streptomycin, 1× non-essential amino acids (NEAA; Thermo Fisher Scientific; Cat. No.: 11140-050) and 1× GlutaMax (Thermo Fisher Scientific; Cat. No.: 35050-061). All cells were cultured at 37°C in a humidified-air 10% CO_2 atmosphere and were verified for the absence of mycoplasma.

Recombinant DNA

The generation of the various recombinant DNA constructs used in this study was done by using standard molecular cloning techniques. The annotated maps and nucleotide sequences of pegRNA expression plasmids S68_pU6.pegRNA$^{EX51.A1.RE}$, BG40_pU6.epegRNA$^{EX51.A1.RE}$, BG42_pU6.epegRNA$^{EX51.7.DEL}$ and BG43_pU6.epegRNA$^{EX51.7.INS}$, BK10_pU6.epegRNA$^{TWIN.PE}$ and the prime editor expression plasmids S65_pCAG.PE.rBGpA and BG50_pCAG.PEmax.rBGpA, are available in pages 1–16 of the **Supplementary Information**. In addition, the oligonucleotides used for the assembly of the various gRNA, pegRNA and epegRNA expression constructs are listed in **Supplementary Table S1**.

DNA transfections

The plasmid DNA transfection screens used to identify functional prime editing reagents were initiated by seeding HEK293T cells at a density of 2.0×10^5 cells per well of 24-well plates (Greiner Bio-One). After overnight incubation, the cells were transfected with the aid of 1 mg ml^{-1} 25 kDa linear polyethyleneimine (PEI, Polysciences) solution (pH 7.4) following the protocol described previously (34). The compositions of the different plasmid transfection reactions are specified in **Supplementary Tables S2** and **S3**. At 3 days post-transfection, the cells were harvested for target-site genotyping analysis.

AdVP production, purification and characterization

The production of AdVP.PE2$^{DMD.INS+1}$, AdVP.PE3$^{DMD.DEL-2}$, AdVP.PE3$^{DMD.INS+1}$ and AdVP.GpNLuc was done in bacteriophage P1 Cre recombinase-expressing PER.tTA.Cre43 cells (30) derived from the adenovirus type 5 *E1*-complementing packaging cell line PER.C6 (35). AdVP.GpNLuc encodes the reporter GpNLuc, a fusion product between EGFP and NanoLuc (36). The PER.tTA.Cre43 cells were seeded at a density of 1.8×10^6 cells per well of 6-well plates (Greiner Bio-One). The next day, the cells were transfected with 6.25 μg of MssI-linearized plasmids BG59_pAdVP.PE2$^{DMD.INS+1}$, BG62_pAdVP.PE3$^{DMD.DEL-2}$, BG63_pAdVP.PE3$^{DMD.INS+1}$, BK17_AdVP.TwinPE$^{\Delta EX51}$ or BJ03_AdVP.GpNLuc with the aid of PEI. After a 6-h incubation, the transfection medium was replaced by fresh medium containing *E1*-deleted helper AdV vector AdV.SRα.LacZ.1.50 (37) at a multiplicity of infection (MOI) of 5 transducing units (TU) per cell. The producer cells were harvested upon the emergence of complete cytopathic effect (CPE) and were then subjected to three cycles of freezing and thawing in liquid N_2 and 37°C water baths, respectively. Cellular debris were subsequently removed by centrifugation for 10 min at 2,000×g. After three rounds of propagation in PER.tTA.Cre43 cells in the presence of helper AdV.SRα.LacZ.1.50, the supernatant was harvested from twenty T175-cm^2 culture flasks (Greiner Bio-One) each containing 2.3×10^7 producer cells. Next, sequential block and continuous CsCl buoyant density ultracentrifugation was performed for purifying the vector particles present in clarified producer-cell supernatants generated after treatments with sodium deoxycholate detergent and DNaseI at 20 μg ml^{-1} (Roche; Cat. No.:10104159001). The purified vector particles were then de-salted by ultrafiltration through Amicon Ultra-15 100K MWCO filters (MerckMillipore; Cat. No:UFC910024).

Restriction enzyme fragment length analysis (RFLA) was used to determine the structural integrity of vector genomes packaged in purified adenoviral capsids. In brief, vector DNA was isolated by using the DNeasy Blood & Tissue Kit (QIAGEN; Cat. No.: 69506) with the recovered vector genomes being subsequently subjected to specific restriction enzyme digestions. The parental and helper plasmids were digested in parallel with the same restriction enzyme to serve as molecular weight references. After agarose gel electrophoresis, the digested fragments were analyzed by using the Gel-Doc XR+ system and the ImageLab (version 6.0.1) software (both from Bio-Rad). The in silico restriction patterns

corresponding to intact vector genomes and respective vector molecular clone plasmids were made with the aid of SnapGene (version 6.0.7) software.

The AdVP transducing titers were determined through quantitative PCR (qPCR) assays following previously detailed procedures (27). In brief, HeLa cells were plated at a density of 8×10^4 cells per well of 24-well plates (Greiner Bio-One). The next day, the cells were transduced with 5 serial 3-fold dilutions of each of the 100-fold diluted purified AdVP preparations. At approximately 24 h post-transduction, total cellular DNA was extracted from transduced cells via the DNeasy Blood & Tissue kit using the manufacturer's instructions. In parallel, 8 serial 10-fold dilutions of a linearized AdVP molecular clone plasmid (1×10^7 genome copies per microliter) was prepared for the generation of standard curves. Next, a qPCR specific for the AdVP DNA packaging signal was carried out on the cellular and standard curve DNA templates by using the CFX Connect Real-Time PCR Detection System (Bio-Rad) with the Bio-Rad CFX Manager (version 3.1) software being applied for data analysis. The primers, cycling conditions and components of qPCR mixtures applied are specified in **Supplementary Tables S4** and **S5**. The genome-editing AdVP MOIs indicated in this study were based on the transducing titers listed in **Supplementary Table S6**.

Muscle cell differentiation assays

Skeletal muscle cell differentiation was initiated by plating human myoblasts in 0.1% (w/v) gelatin (Sigma-Aldrich; Cat. No.: G1393) coated wells. After reaching full confluency, myoblasts were incubated in differentiation medium consisting of phenol red-free DMEM (Thermo Fisher Scientific; Cat. No.: 11880-028), 100 U ml^{-1} penicillin/streptomycin, 100 µg ml^{-1} human holo-transferrin (Sigma-Aldrich; Cat. No.: T0665) and 10 µg ml^{-1} human insulin (Sigma-Aldrich; Cat. No.: I9278). At approximately four days post-differentiation, the cultures of post-mitotic myotubes were processed for downstream analyses.

The DMD iPSCs were differentiated into beating cardiomyocytes following the protocol for cardiac lineage specification based on a stepwise supplementation of iPSC medium with specific small molecules as detailed elsewhere (38). Briefly, DMD iPSCs cultured in mTeSR medium supplemented with RevitaCell (1:200) were seeded in wells of 12-well plates coated with Matrigel at a density of $3x10^5$ cells per well. At 24 h after seeding, the culture medium was replaced by modified LI-BPEL (mBEL) medium supplemented with 5 M CHIR 99021 (Axon Medchem; Cat. No.: Axon1386) and, 48 h later, this medium was replenished by mBEL medium supplemented with 5 M XAV 939 (Tocris; Cat. No.: 3748/10) and 0.25 M IWPL6 (AbMole; Cat. No.: M2781). After two additional days, the medium was again replenished with mBEL medium, this time supplemented with Insulin-Transferrin-Selenium Ethanolamine (ITS-X) (1:1000) (Thermo Fisher; Cat. No.: 51500-056). At this stage, the cell differentiation medium was replenished every 2 days with areas of beating cardiomyocytes starting to emerge from day 10 onwards (https://doi.org/10.6084/m9.figshare.24869136). After 21 days under cardiomyogenic differentiation conditions, the cells were dissociated and processed for prime editing experiments using AdVP delivery.

Transduction experiments

The transduction of myoblasts and hMSCs was carried out as follows. Approximately 16- to 18-h prior to transduction, human myoblasts and hMSCs were seeded in 24-well plates at a density of 8×10^4 and 1×10^5 cells per well, respectively. The next day, these cells were transduced with AdVPs at the MOIs specified in the corresponding figures. Three days after transduction, the cells were transferred to a wells of 6-well plates and were sub-cultured for another seven days. Afterwards, genomic DNA from the transduced cells was isolated by using the DNeasy Blood & Tissue Kit for assessing prime-editing activities. Prime edits and bystander events in the form of indels and pegRNA scaffold-derived insertions were quantified and characterized by next-generation deep sequencing.

Experiments designed for testing the stacking of prime-editing events in target cell populations by successive transduction rounds were carried out in wild-type human myoblasts as follows. One day after seeding in 6-well plates at a density of 5×10^5 cells per well, the wild-type myoblasts were transduced with AdVP.$PE2^{DMD.INS+1}$ at an MOI of 50 TU $cell^{-1}$. After overnight incubation, the medium was substituted by fresh medium and, at three days post-transduction, fractions of myoblast suspensions were harvested for genomic DNA extraction and the remaining cell suspension bulks were seeded for a second AdVP transduction round. The same procedures were applied for the third and final AdVP

transduction round. Finally, genomic DNA samples, collected via the DNeasy Blood & Tissue Kit, were subjected to high-throughput next-generation sequencing analysis for quantifying and characterizing prime-editing events at DMD target alleles.

The comparison of prime editing activities in cycling wild-type myoblasts versus post-mitotic myotubes was initiated by seeding 5×10^4 and 2×10^5 wild-type myoblasts in wells of 24-well plates. The next day, the former cells were transduced with AdVPs at 50 TU cell^{-1} and 100 TU cell^{-1}, and the latter cells were exposed to mitogen-poor differentiation medium consisting of phenol red-free DMEM, 100 U ml^{-1} penicillin/streptomycin, 100 µg ml^{-1} human holo-transferrin and 10 µg ml^{-1} human insulin. Two days after differentiation initiation, the myotubes were treated with AdVPs at 50 TU cell^{-1} and 100 TU cell^{-1}. All the culture vessels used in this study for myotube culturing are pre-coated with 0.1% gelatin solution. AdVP-treated myoblasts and myotubes were harvested for western blotting and genomic DNA extraction at 2 and 3 days post-transduction, respectively.

Transduction experiments in DMD iPSC-derived cardiomyocytes were initiated by seeding the differentiated cardiomyocytes in wells of 96-, 48- and 24-well plates at a density of 1×10^5 cells per cm^2 in mBEL medium supplemented with ITS-X (1:1000) and a 1:200 dilution of RevitaCell. In particular, after 20 days of differentiation, the cardiomyocytes were first dissociated by incubation for 5 min at 37°C in 1× TrypLE Select (Thermo Fisher; Cat. No.: A1217701). The resulting cell suspensions were then seeded in the appropriate multi-well plates previously coated with Matrigel. Three days later, the medium was replaced by the appropriate amount of medium containing AdVPs at different MOIs and 1.5 h later, the culture medium was replenished once again. At 4 days post-transduction, the cardiomyocytes were harvested for genomic DNA analysis and, at 8 days post-transduction, they were collected for reverse transcription-qPCR (RT-qPCR) and immunofluorescence microscopy analyses. The AdVP MOI ranges applied to the myoblasts derived from different donors and to the other myogenic cell types tested, i.e. mesenchymal stem cells and iPSC-derived cardiomyocytes, did not lead to noticeable cytotoxic effects in the transduced cells.

On-target and off-target sites genotyping assays

Prime editing activities in HEK293T cells transfected with prime editing constructs were assessed through the analysis of Sanger sequencing chromatogram peaks by using the Inference of CRISPR Edits (ICE) or Tracking of Indels by Decomposition (TIDE) software packages (39,40). In brief, genomic DNA derived from treated and untreated samples was extracted with the DNeasy Blood & Tissue kit following the manufacturer's recommendations. Next, the target sites were amplified by using Phusion High-Fidelity Polymerase (Thermo Fisher Scientific; Cat. No.: #F-530L). The primer sequences, PCR mixture compositions and cycling conditions applied are listed in **Supplementary Tables S7** and **S8**. The resulting amplicons, purified by using the QIAEX II Gel Extraction Kit (QIAGEN; Cat. No.: 20021) or Mag-Bind XP beads, were then subjected to Sanger sequencing with the amplicon chromatograms derived from treated and untreated samples serving as input for TIDE or ICE analyses (39,40).

The frequencies of AdVP-induced prime edits and bystander events in the form of indels and pegRNA scaffold-derived insertions were quantified and characterized by high-throughput NGS analysis following a protocol detailed elsewhere (34). In brief, the DNeasy Blood & Tissue Kit was used to extract genomic DNA from mock-transduced myoblasts and hMSCs or from AdVP-transduced myoblasts and hMSCs and, subsequently, the extracted DNA was subjected to gene-specific PCR amplification using the Phusion High-Fidelity Polymerase. The resulting amplicons were purified with AMPure XP beads (Beckman Coulter; Cat. No.: A63881) and then were subjected to barcoding PCR. The primers, cycling parameters and PCR mixtures used for the preparation of gene-specific and barcoded amplicons are indicated in **Supplementary Tables S9-S13**. Gene-specific amplicons corresponded to the *DMD* target DNA and to the first three top-ranked candidate off-target sites for the spacer of the DMD-targeting epegRNAs. These candidate off-target sites map at an intergenic sequence of the *SLITRK5-LINC00397* locus and at intronic sequences of the *STRIP1* and *VGLL4* genes and were identified by using the CRISPOR algorithm (41). The concentrations of barcoded amplicons were measured by Qubit2.0 fluorometer (Invitrogen) with the Qubit dsDNA HS assay kit (Invitrogen; Cat. No.: Q32854) and the quality of barcoded amplicon library was assessed by 2100 Bioanalyzer system (Agilent). Finally, amplicons were pooled in equal molar ratios and subjected to next-generation Illumina MiSeq deep sequencing for obtaining 50,000 paired-end reads on a per sample basis. CRISPResso2 software (42)

was applied for data analyses after demultiplexing of the paired-end MiSeq raw reads (R1 and R2 fastq files). The quality control of the circa 50,000 paired-end reads per sample and the scripts applied for the CRISPResso2 analyses are available in the **Supplementary Information**.

Reverse transcription-qPCR

The quantification of DMD mRNA levels in unedited and AdVP-edited myoblasts was done by reverse transcription-qPCR (RT-qPCR) as follows. First, differentiation was induced in mock- and AdVP-transduced myogenic progenitor cells and, upon the formation of post-mitotic myotubes or cardiomyocytes, RNA was extracted by using the NucleoSpin RNA Kit following the manufacturer's instructions (Macherey Nagel; Cat. No.: 740955). The concentration of isolated RNA was determined by a Nanodrop apparatus and then equal amounts of RNA was reverse transcribed with the aid of the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific; Cat. No.: K1691). In brief, 1000 ng of RNA was incubated with 0.5 µl of 100 µM random hexamer primers and 0.5 µl of 100 µM Oligo(dT)$_{18}$ primers in 12 µl reaction volumes at 65°C for 5 min followed by an 2-min incubation at 4°C. Subsequently, 1 µl of 20 U µl^{-1} RiboLock RNase Inhibitor, 1 µl of 200 U µl^{-1} RevertAid H Minus M-MuLV Reverse Transcriptase, 2 µl of 10 mM dNTP Mix and 4 µl of 5× Reaction Buffer, were directly added to each sample and the resulting mixtures were incubated at 25°C for 5 min followed by an 1-h incubation at 42°C. Afterwards, the reverse transcriptase was inactivated by heating the samples at 70°C for 5 min. The synthesized cDNA templates were then diluted 5-fold in nuclease-free water and 1 µl of the diluted cDNA template was used for qPCR amplification targeting DMD sequences with the aid of iQ™ SYBR® Green Supermix (Bio-Rad; Cat. No.: L010171C) and the primers indicated in **Supplementary Table S4**. In addition, target information, qPCR mixture components, cycling conditions and amplicon sizes are specified in **Supplementary Tables S4** and **S5**, respectively. Housekeeping *GAPDH* transcripts served as internal control target templates for gene expression normalization. The specificity of each primer pair was predicted by in silico BLAST screens and then validated with qPCR melting profile. The qPCR signal outputs were detected with the CFX Connect Real-Time PCR Detection System (Bio-Rad) and then analyzed by using the $2^{-\Delta\Delta Ct}$ method to determine the relative expression levels. Statistical analyses were done with the GraphPad Prism software (version 9.3.1).

Western blotting

Myotubes differentiated from mock- and AdVP-transduced myoblasts were lysed with Laemmli buffer consisting of 8.0% glycerol, 3% sodium dodecyl sulfate (SDS) and 200 mM Tris-HCl (pH 6.8). Protein concentrations were determined with the DC™ protein assay Kit (Bio-Rad; Cat. No.: 5000111) according to the manufacturer's protocol. Afterwards, equal amounts of proteins extracted from experimental and control samples and a dose-range of proteins extracted from healthy donor myoblasts were loaded and separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The loaded protein amounts and gel resolution used are specified in the corresponding figure legends. Subsequently, the resolved proteins were transferred onto 0.45-µm polyvinylidene difluoride (PVDF) membranes (Merck Millipore; Cat. No.: IPVH00010) at 60 V for 24 h, after which the membranes were blocked with 5% non-fat dry milk dissolved in Tris-buffered saline (TBS) with 0.1% Tween 20 (TBST) at room temperature (RT) for at least 1 h. Next, the membranes were incubated overnight at 4°C with the respective primary antibodies, i.e., anti-dystrophin (1:500 dilution; Abcam; Cat. No.: ab15277), anti-myosin heavy chain (1:500 dilution; Sigma-Aldrich; Cat. No.: M4276), anti-vinculin (1:1000 dilution; Sigma-Aldrich; Cat. No.: V9131), anti-GAPDH (1:1000 dilution; Merck Millipore; Cat. No.: MAB374) or anti-α/β-Tubulin (1:1000 dilution; Cell Signaling Technology; Cat. No.: 2148). After thrice washes with TBST, the membranes were probed with the appropriate secondary antibodies, i.e., anti-mouse IgG (1:5000 dilution; Sigma-Aldrich; Cat. No.: NA931V) or anti-rabbit (1:1000 dilution; Cell Signaling; Cat. No.: 7074S) at RT for 2 h. Finally, signal detection was carried out by using Clarity™ Western ECL Substrate (Bio-Rad; Cat. No.: 1705060) together with the ChemiDoc Imaging System (Bio-Rad; Cat. No.: 17001402).

Immunofluorescence microscopy analyses

Dystrophin expression in and differentiation capacity of human myoblasts subjected to AdVP-based *DMD* prime editing was assessed by immunofluorescence staining and confocal microscopy. In brief, DMD.D48–50 myoblasts (AB1098) edited via transduction with prime-editing AdVPs were incubated in myogenic differentiation medium for approximately four days and subsequently fixed with 4% paraformaldehyde (PFA), permeabilized with 0.5% Triton X-100 in TBS (50 mM Tris-HCl pH 7.5 with 100 mM NaCl) and blocked in TBS containing 0.1% Triton X-100, 2% bovine serum albumin (BSA) and

0.1% sodium azide. Cultures of unedited DMD.D48–50 myoblasts (AB1098) were equally processed in parallel Next, experimental and control specimens were incubated overnight at 4°C with the appropriate primary antibodies indicated in **Supplementary Table S14** and, after three 10-min washes with TBS, the target antigens were probed with fluorochrome-conjugated secondary antibodies (**Supplementary Table S14**). Afterwards, the specimens were mounted in ProLong Gold Antifade Mounting reagent containing DAPI (Thermo Fisher Scientific; Cat. No.: P36931). Finally, images were acquired with the aid of an upright Leica SP8 confocal microscope equipped with Leica hybrid detectors HyD and analysed with the LAS X software (Leica Microsystems). Dystrophin expression in DMD iPSC-derived cardiomyocytes that were mock-transduced or AdVP-transduced was also assessed through immunofluorescence microscopy analysis. In brief, at 8t days post-transduction, cells previously seeded in wells of 96-well plates were subjected to the above-described staining protocol except that they were incubated overnight at 4°C in blocking solution. Next, the cells exposed and not exposed to AdVP transduction were sequentially incubated for 2 h at RT with the C-terminal-specific anti-Dystrophin antibody ab15277 (Abcam), diluted 1:100, and then with the Alexa Fluor 488 goat anti-rabbit IgG (H+L) secondary antibody diluted 1:500 in TBS containing 2% BSA. The cell nuclei were stained by incubation with Hoechst 33342 (Invitrogen; Cat. No.: H3570) diluted 1:1000 in PBS for 10 min at RT. Finally, images were acquired by using a AF6000 LX microscope and subsequently analyzed with the aid of the ImageJ software (NIH, US National Institutes of Health).

Proximity ligation assays

Besides dual color fluorescence microscopy, the colocalization of dystrophin and β-dystroglycan was detected by using a proximity ligation assay (PLA). In brief, DMD.D48–50 myoblasts (AB1098 and 6594) edited via transduction with prime-editing AdVPs, were seeded in the wells of a 24-well plate containing coverslips pre-coated with 0.1% gelatin. Myogenic differentiation was triggered once the cells reached full confluence and, at approximately four days post-differentiation, the cells were fixed with 4% PFA for 10 min and permeabilized with 0.5% Triton X-100 in PBS for 5 min at RT. Subsequently, the specimens were blocked with Duolink® Blocking Solution (Sigma-Aldrich; Cat. No.: DUO82007) for 1 h at 37°C in a heated humidity chamber and were then incubated overnight at 4°C with primary antibodies against the C-terminus of dystrophin (1:100 dilution; Abcam; Cat. No.: ab15277) and β-dystroglycan (1:100 dilution; Santa Cruz Biotechnologies; Cat. No.: sc-33702) diluted in Duolink® Antibody Diluent (Sigma-Aldrich; Cat. No.: DUO82008). After three washes with Duolink® Wash Buffer A (Sigma–Aldrich; Cat. No.: DUO82046), the specimens were exposed to secondary antibodies conjugated to Duolink® PLUS and MINUS PLA probes (Sigma–Aldrich; DUO92001 and DUO92005) in Duolink® Antibody Diluent at a 1:10 dilution for 1 h at 37°C in a heated humidity chamber. Afterwards, the coverslips were washed twice with Duolink® Wash Buffer A and then exposed to Duolink® Ligase (Sigma–Aldrich; Cat. No.: DUO82027) in 1× Duolink® Ligation Buffer (Sigma–Aldrich; Cat. No.: DUO82009) at a 1:40 dilution for 30 min at 37°C in a heated humidity chamber. After two 5-min washes, the specimens were incubated with Duolink® Polymerase (Sigma–Aldrich; Cat. No.: DUO82028) in 1× Duolink® Amplification Buffer (Sigma–Aldrich; Cat. No.: DUO82011) at a 1:80 dilution for 100 min at 37°C in a heated humidity chamber. Next, the specimens were washed twice with 1× Duolink® Wash Buffer B (Sigma–Aldrich; Cat. No.: DUO82048), followed by a brief wash with 0.01× Duolink® Wash Buffer B for 1 min. Finally, the samples were mounted in VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories; Cat. No.: H-1200). Images were acquired by using a Leica SP8 confocal microscope equipped with Leica hybrid detectors HyD and analyzed with the LAS X software (Leica Microsystems).

Flow cytometry

The expression levels of CAR and CD46 on myoblasts from different donors were determined by using a BD LSR II flow cytometer (BD Biosciences). In brief, cells were harvested and washed with PBS supplemented with 1% BSA. Next, 1×10^5 myoblasts were resuspended in 100 µl of ice-cold PBS containing 2% BSA and then stained with 5 µl of a FITC-conjugated anti-CAR antibody (Santa Cruz Biotechnologies; Cat. No.: SC373791) or 5 µl of a PE-conjugated anti-CD46 antibody (Thermo Fisher Scientific; Cat. No.:12-0469-42) for 30 min on ice in the dark. After thrice washing with ice-cold PBS supplemented with 1% BSA, the cells were resuspended in 300 µl of PBS containing 0.5% BSA and 2 mM EDTA (pH=8.0). Myoblasts incubated with equal amounts of a FITC-conjugated IgG2b isotype control antibody (Santa Cruz Biotechnologies; Cat. No.:SC2857) or a PE-conjugated IgG1 kappa isotype control antibody (Thermo Fisher Scientific; Cat. No.:12-4714-82) served as negative controls to

establish the thresholds for background fluorescence. At least 10,000 viable single cells were acquired per sample. Data were analyzed with the aid of the FlowJo software (Tree Star; version 10.5.0).

Statistical analyses

Statistical analyses were performed with the aid of the GraphPad Prism software (version 9.3.1) on datasets derived from a minimum of three biological replicates. Two-tailed unpaired Student's t tests were carried out to analyze whether there were any statistically significant differences between two unrelated groups, whereas analysis of variance (ANOVA) was used for determining the statistical significance of three or more independent groups. One-way ANOVA and two-way ANOVA were performed on the datasets with one independent factor and two independent factors, respectively, and whenever there was a statistical significance, multiple comparison tests were followed. Dunnett's multiple comparison tests were applied for comparing each mean to a control mean, while Tukey's multiple comparison tests were used to compare each mean with each other mean. Details on statistical parameters and tests used in each experiment are specified in the respective figure legends. P values lower than 0.05 were considered to be statistically significant.

RESULTS

We started by performing transient transfection experiments in HEK293T cells for assessing the original PE2 prime editor versus the optimized PEmax variant mixed with regular pegRNAs or end-protected epegRNAs either alone (PE2 setups) or together with auxiliary gRNAs (PE3 setups). Both types of pegRNAs were designed for installing frameshifting 1-bp insertions or 2-bp deletions at *DMD* exon 51 together with 1-bp substitutions for blunting target site re-engagement through protospacer adjacent motif (PAM) elimination. Genotyping assays confirmed that epegRNAs and PEmax can foster chromosomal DNA editing (**Supplementary Figure 1**) and identified combinations of prime editing reagents designed for disrupting and restoring the *DMD* reading frame in cells with wild-type and DMD-causing genotypes, respectively (**Supplementary Figure 2**). Thus, based on these DNA transfection screens, constructs encoding optimized prime editing reagents were selected for packaging in adenoviral capsids resulting in AdVP.$^{PE2DMD.INS+1}$, for testing PE2-mediated 1-bp insertions, or in AdVP.PE3$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.DEL-2}$, for testing PE3-mediated 1-bp insertions and 2-bp deletions, respectively (**Figure 1A**). Of notice, instead of prototypic adenovirus type-5 fibers, these AdVPs were endowed with type-50 fibers to, via CD46-binding, bypass the absence of the coxsackievirus and adenovirus receptor (CAR) on human myogenic cell types, namely, mesenchymal stem cells and bona fide muscle progenitors (43,44). The absence of CAR and the presence of CD46 on myoblasts derived from healthy and DMD donors was confirmed through flow cytometry analysis (**Supplementary Figure 3**). Moreover, transduction experiments with a reporter AdVP vector displaying type-50 fibers, established efficient transduction of CAR-negative muscle progenitors by CD46-binding vector particles (**Supplementary Figure 4**).

Stocks of AdVP.PE2$^{DMD.INS+1}$, AdVP.PE3$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.DEL-2}$ were produced to similar high titers, i.e., 1.80×10^{10} transducing units per ml (TU ml^{-1}), 1.11×10^{10} TU ml^{-1} and 1.76×10^{10} TU ml^{-1}, respectively, and contained structurally intact vector genomes with evidence neither for rearranged nor truncated species (**Supplementary Figure 5**). Importantly, transduction experiments using prime editing CD46-targeting AdVPs in human myoblasts originated from three different DMD patients with intragenic deletions revealed a clear AdVP dose-dependent increase in the frequencies of *DMD* edition regardless of the construct used as determined by high-throughput deep sequencing (**Figure 1B** and **Supplementary Figure 6**) and inference of CRISPR edits (ICE) analyses (39) (**Supplementary Figure 7**). DMD edition upon AdVP delivery of PE3 machineries was superior to that resulting from PE2 transfer in human myoblasts (**Figure 1B** and **Supplementary Figure 6**) and, even more so, in human mesenchymal stem cells (hMSCs) used here as an independent cell type with myogenic capacity (**Supplementary Figure 8**). As expected, when compared to the PE3 machineries, PE2 led to lower frequencies of byproducts in the form of imprecise indels and epegRNA scaffold-derived insertions (**Figure 1B**, **Supplementary Figure 6** and **Supplementary Figure 8**). Interestingly, although precision indexes corresponding to edit-to-byproduct ratios were highest for PE2 complexes in human myoblasts (**Figure 1C** and **Supplementary Figure 6B**), these indexes were similar amongst PE2 and PE3 complexes in hMSCs (**Supplementary Figure 8B**). Considering that human myoblasts and hMSCs are transduced equally well by CD46-binding AdVPs (27), these data support the proposition that cell type-specific determinants, namely, complement of DNA repair factors (8,22), cell-cycle activity (22,27)

and/or target chromatin context (45), in addition to affecting the efficiency of prime editing, can also have a bearing on its ultimate product purity.

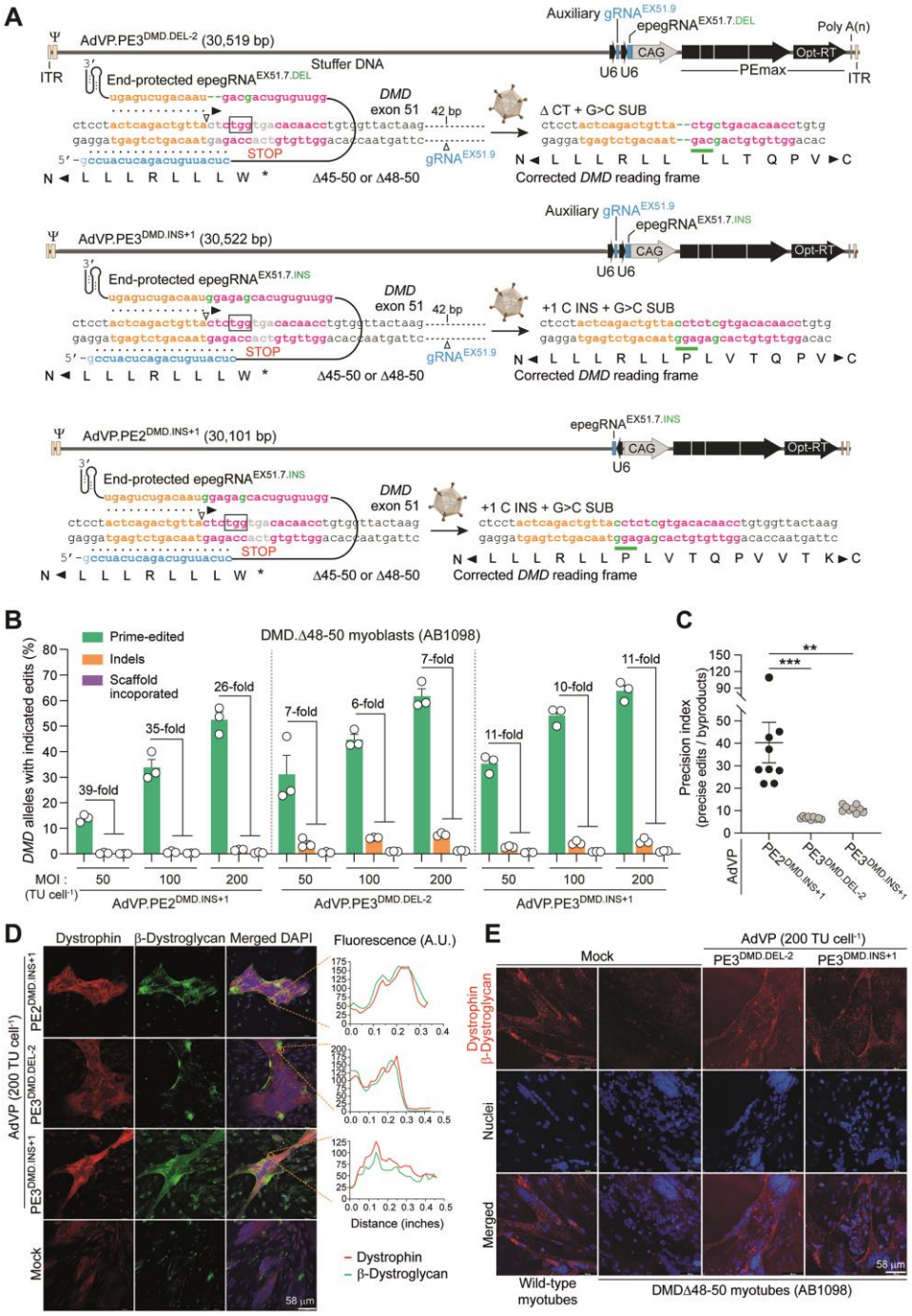


Figure 1. Gene correction through AdVP-based prime editing in *DMD* defective myoblasts. (A) Genome structures of AdVPs assembled for *DMD* prime editing. *DMD* target sequences before and after prime editing are depicted. The hybrid CAG promoter drives PEmax expression whilst the human U6 promoter controls the synthesis of the indicated epegRNAs and gRNAs. Spacer, primer binding site (PBS) and reverse transcriptase template (RTT) sequences of epegRNA are marked in cyan, orange and magenta, respectively, with encoded and installed edits labelled in green. Protospacer adjacent motifs (NGG) are boxed, and nicking positions are marked by open arrowheads. ITR and Ψ, adenovirus type-5 *cis*-acting inverted terminal repeats and packaging signal, respectively. (**B**) AdVP-based prime editing in DMD myoblasts. Human myoblasts with a Δ48-50 genotype (DMD.Δ48-50) were transduced with different multiplicities-of-infection (MOI) of AdVP.PE2$^{DMD.INS+1}$, AdVP.PE3$^{DMD.DEL-2}$ and AdVP.PE3$^{DMD.INS+1}$; TU cell^{-1}, transducing units per cell. Prime edits and unwarranted byproducts (i.e., indels and scaffold-derived insertions) were quantified by next-generation deep sequencing at 10 days post-transduction (50,000 paired-end reads per sample). Bars and error bars denote mean ± SEM, respectively, of three biological replicates. (**C**) Prime-editing precision indexes upon AdVP transduction. Precision indexes corresponding to the cumulative ratios of precise edits to byproducts frequencies measured in AdVP-transduced myoblasts DMD.Δ48-50 (AB1098) are plotted as mean ± SEM of the independent datapoints. Significances were calculated with one-way ANOVA followed by Dunnett's multiple comparison tests; ***0.0001<*P*<0.001,

0.001<*P*<0.01. (D**) Detection of dystrophin and β-dystroglycan in DMD.Δ48-50 (AB1098) muscle cells prime-edited using AdVPs. Dual-color immunofluorescence microscopy for dystrophin and β-dystroglycan was done on myotubes differentiated from DMD.Δ48-50 myoblasts transduced with the indicated DMD prime-editing AdVPs. Co-localization of dystrophin and β-dystroglycan at the plasma membrane of prime-edited DMD myotubes was assessed by image merging and dystrophin- plus β-dystroglycan-specific fluorescence signal measurements (boxed areas). Nuclei are labeled with DAPI in the merged images. (**E**) Dystrophin-β-dystroglycan interaction analysis in DMD muscle cells prime-edited using AdVPs. Proximity ligation assay detection of endogenous dystrophin-β-dystroglycan interactions was carried out on myotubes differentiated from DMD.Δ48–50 myoblasts transduced with the indicated DMD prime-editing AdVPs (red foci). Healthy donor (wild-type) and untreated DMD patient-derived myotubes served as positive and negative controls, respectively. Nuclei were labelled by DAPI staining.

Next, we sought to assess *DMD* gene expression upon myogenic differentiation of AdVP-edited DMD myoblasts. Firstly, myogenic differentiation capabilities amongst untreated and AdVP-treated myoblasts were not overtly different as probed via immunofluorescence microscopy directed at late muscle-specific markers, i.e., skeletal fast-twitch myosin heavy chain and sarcomeric a-actinin (**Supplementary Figure 9**). Secondly, consistent with *DMD* reading frame resetting, evidence for *de novo* expression of Becker-like dystrophin transcripts in differentiated myotubes was obtained via a combination of RT-qPCR assays targeting edited and unedited mRNA sequences (**Supplementary Figure 10A** and **10B**, respectively). Indeed, the latter RT-qPCR assays specific for sequences upstream and downstream of target exon 51, consistently measured a significant increase in *DMD* mRNA transcript levels in myotubes differentiated from muscle progenitors initially transduced with AdVP.PE3$^{DMD.INS+1}$ or AdVP.PE3$^{DMD.DEL-2}$ (**Supplementary Figure 10B**). Additional RT-qPCR assays specific for distal mRNA sequences encoding dystrophin C-terminal domains confirmed that, when compared to mock controls, AdVP transductions resulted in higher amounts of *DMD* mRNA transcripts in differentiated muscle cells (Supplementary Figure 11). These data indicate that DMD reading frame correction and premature stop codon elimination in prime-edited muscle cells leads to the stabilization of *DMD* transcripts presumably via an interference with otherwise operative nonsense-mediated RNA decay processes. Finally, the expression of Becker-like dystrophins was confirmed at the protein level by immunofluorescence microscopy and western blot analyses (**Figure 1D** and **Supplementary Figure 12**, respectively).

Dystrophin links the internal cytoskeleton to the DGC at the sarcolemma of striated muscle cells via binding to the transmembrane protein β-dystroglycan. In the absence of functional dystrophin molecules, and alike to other DGC proteins, β-dystroglycan presents a shorter half-life and mostly vacates the plasma membrane (46). Importantly, evidence for the stabilization and proper relocation of β-dystroglycan to the plasmalemma of differentiated AdVP-edited DMD muscle cells was provided by dual-colour confocal microscopy analysis of dystrophin and β-dystroglycan (**Figure 1D**). Moreover, proximity ligation assays (**Supplementary Figure 13**), besides independently confirming *de novo* assembly of Becker-like dystrophins in prime-edited myotubes, demonstrated the capacity of these shortened dystrophins to locally associate with β-dystroglycan (**Figure 1E** and **Supplementary Figure 14**).

Prime editing depends firstly on the complementarity of target DNA to spacer and PBS sequences in the pegRNA and, secondly, on the complementarity of the reverse transcribed template to the target sequence. As a result of these multitier hybridization requirements, prime editing at off-target positions is significantly rarer than Cas9:gRNA-induced off-target mutations in that the latter only require a single spacer-protospacer hybridization interrogation step. Nonetheless, as initially shown in a bacteriophage replication system (47), nicks can in principle also lead to mutagenic DSBs in mammalian cells if a replication fork advances through them and collapses. Indeed, earlier experiments from our laboratory using unbiased high-throughput genome-wide translocation sequencing (HTGTS) revelated that, albeit at low frequencies, Cas9 nickases do trigger chromosomal break-derived translocations involving gRNA off-target positions (34,48). Moreover, in view of the therapeutic relevance of DMD-targeting prime editing reagents, we set out to probe their specificity directly in a target cell type-of-interest (i.e., human myoblasts) by using AdVP.PE2$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.INS+1}$ coupled to next-generation sequencing (NGS) analysis at the first three top-ranked candidate off-target sites, i.e., *SLITRK5*, *STRIP1* and *VGLL4* (**Figure 2A**). To increase the stringency of these genotyping assays, parallel cultures of human myoblasts were also simultaneously treated with each of the prime-editing AdVPs and a second-generation adenovector encoding the *S. pyogenes* Cas9 nuclease. These Cas9 nuclease spiking experiments maximize the chance of detecting off-target genomic modifications if the resulting Cas9:epegRNA complexes turn out to productively engage off-target sequences. NGS reads

corresponding to modified off-target sequences were mostly within background levels in cells exposed exclusively to PE2 or PE3 complexes. Importantly, in cells subjected to prime editor and Cas9 activities, modified off-target sequences were within or slightly above background levels especially at *STRIP1* where a single spacer-protospacer mismatch is identified (**Figure 2A** and **Supplementary Figure 15**). As expected, NGS reads corresponding to modified and to modified plus prime-edited target alleles (**Figure 2B**, light pink and orange sectors, respectively) were substantially higher in cells treated with Cas9 than in cells subjected exclusively to PE2 or PE3 activities (**Figure 2B**). This increase in complex on-target modifications was especially noticeable in myoblasts transduced with AdVP.PE3$^{DMD.INS+1}$, presumably due to the combined effects of coupling Cas9 to epegRNA and gRNA molecules (**Figure 2B**). Of notice, the presence of Cas9 also led to a substantial increase in the amounts of scaffold-derived indels (**Figure 2B**, light red sectors). Taken together, these data established epegRNA$^{EX51.7.INS}$ as having high specificity for DMD exon 51 while confirming the poor genotoxicity of prime editors in general (**Figure 2**).

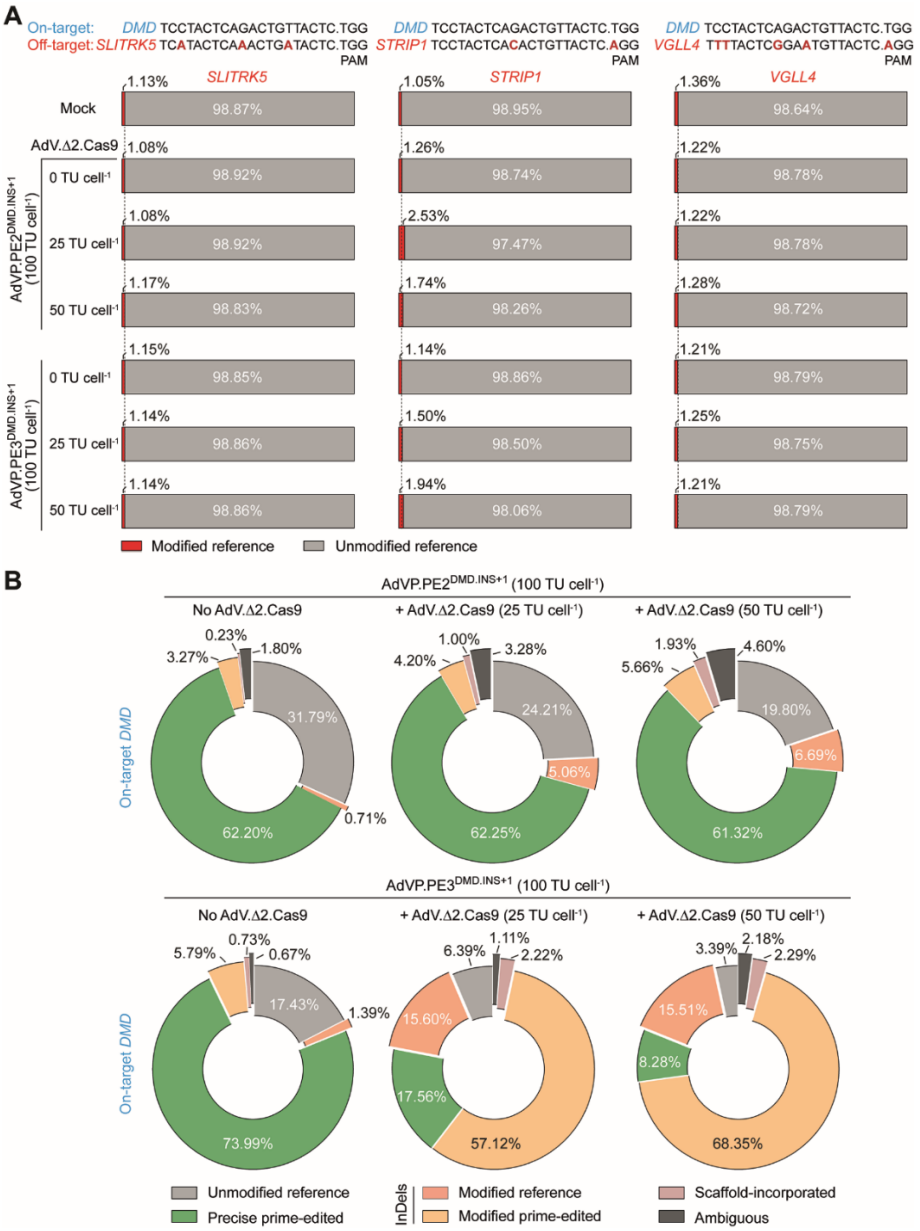


Figure 2. Assessing the specificity of *DMD* prime-editing reagents upon AdVP transduction. (A) Probing off-target activities of *DMD* prime-editing reagents. Human myoblasts DMD.Δ48-50 (AB1098) were individually transduced with AdVP.PE2$^{DMD.INS+1}$ or AdVP.PE3$^{DMD.INS+1}$ at 100 TU cell^{-1} or mixed with Cas9 nuclease-encoding vector AdV.Δ2.Cas9 at 25 TU cell^{-1} and 50 TU cell^{-1}. Genomic modifications at the top-ranked candidate off-target sites *SLITRK5*, *STRIP1* and *VGLL4* were assessed at 3 days post-transduction through NGS analysis (50,000 paired-end reads per sample). These top-ranked candidate off-target sites map at an intergenic sequence in the *SLITRK5-LINC00397* locus and at intronic sequences in the *STRIP1* and *VGLL4* genes. Nucleotide mismatches between target and off-target sites are marked in red. Mock-transduced myoblasts DMD.Δ48-50 (AB1098) provided

for negative controls. Frequencies of NGS reads corresponding to modified and unmodified off-target sequences are shown as red and grey bars, respectively. **(B)** Characterization of target *DMD* gene modifications. Genomic modifications at the *DMD* target region were also determined at 3 days post-transduction through NGS analysis (50,000 paired-end reads per sample). The different types of *DMD* gene modifications identified are indicated and distributed in the part-to-a-whole donut charts.

Nicking-based prime editing is a more predicate and less mutagenic procedure for achieving targeted gene knockouts than NHEJ-based genome editing involving CRISPR nuclease delivery and ensuing DSB formation. Hence, to complement the previous *DMD* gene correction experiments, we advanced to testing AdVP-based prime editing for establishing targeted *DMD* gene knockouts by transducing wild-type myoblasts with AdVP.PE2$^{DMD.INS+1}$ or AdVP.PE3$^{DMD.DEL-2}$ (**Figure 3A**). Interestingly, in these myoblasts, *DMD* editing levels induced by PE3 components were robust but not superior to those triggered by PE2 components (**Figure 3B**). This outcome combined with the rarity of PE2-derived indels resulted in a particularly favourable precision index for the PE2 system in these cells (**Figure 3C**). Finally, the efficient installation of frameshifting 1-bp insertions and 2-bp deletions at wild-type *DMD* alleles by AdVP.PE2$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.DEL-2}$, respectively, correlated with robust gene knockout levels in transduced cells as assessed through RT-qPCR and western blot analyses (**Figure 3D** and **3E**, respectively).

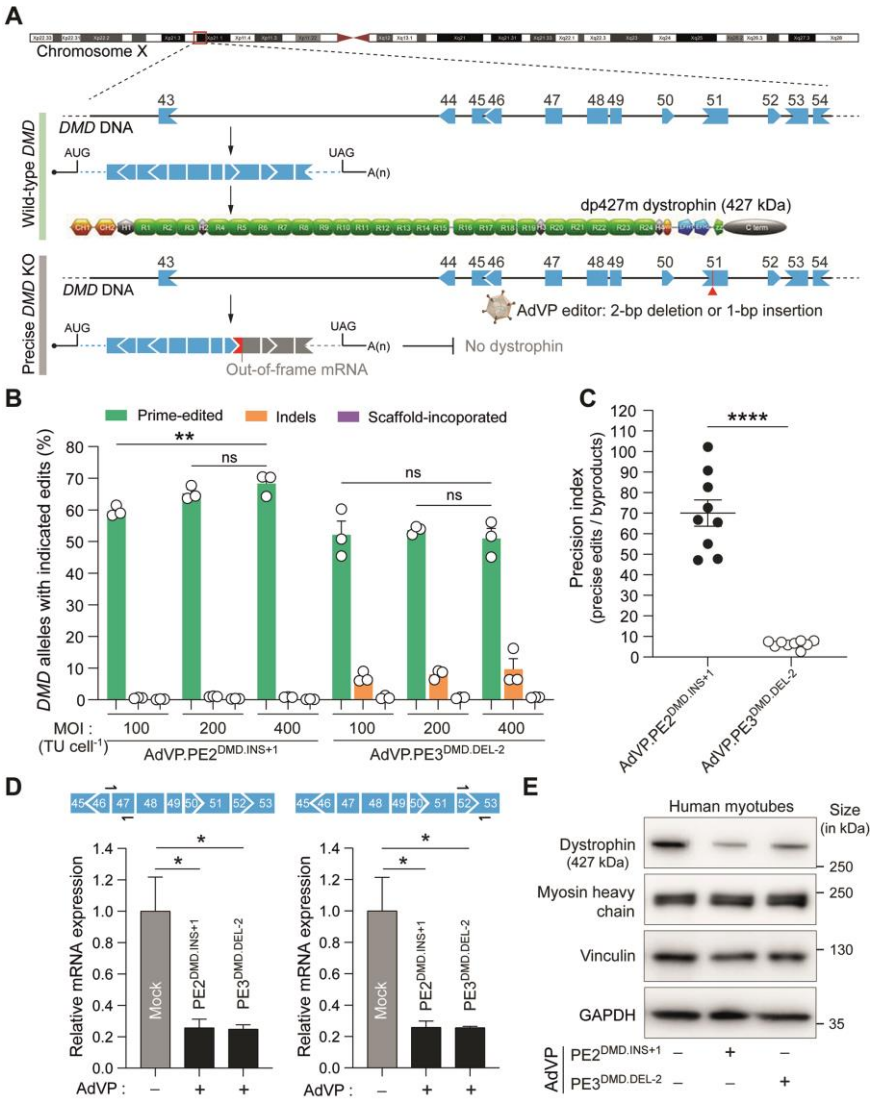


Figure 3. Gene knocking out through AdVP-based prime editing. (A) *DMD* gene knockouts using AdVP-based prime editing in human muscle cells. The synthesis and splicing of transcript isoform Dp427m leads to the assembly of 14-kb mature transcripts coding for 427-kDa dystrophin molecules whose amino and carboxy termini flank a long spectrin-like repeat region and bind to, respectively, F-actin in the cytoskeleton and dystrophin-glycoprotein complexes at the sarcolemma. In wild-type muscle cells, the installation of 2-bp deletions or 1-bp insertions within *DMD* exon 51 upon AdVP-based prime editing results in reading frame disruption and ensuing dystrophin knockout in differentiated muscle cells. (**B**) AdVP-based prime editing in wild-type myoblasts. Human myoblasts with a regular *DMD* genotype were transduced with different multiplicities-of-infection (MOI) of AdVP.PE2$^{DMD.INS+1}$, and AdVP.PE3$^{DMD.DEL-2}$; TU cell^{-1}, transducing units per cell. Prime edits and unwarranted bystander events (i.e.,

indels and scaffold-derived insertions) were measured by high-throughput NGS at 10 days post-transduction (50,000 paired-end reads per sample). Bars and error bars correspond to mean ± SEM, respectively, of three biological replicates. Significances for the indicated datasets were calculated with two-way ANOVA followed by Tukey's multiple comparison tests; **0.001<*P*<0.01; *P*>0.05 was considered non-significant (ns). (**C**) Prime-editing precision indexes upon AdVP transduction. Precision indexes corresponding to the cumulative ratios of precise edits to bystander event frequencies measured in AdVP-transduced wild-type myoblasts are plotted as mean ± SEM of the independent datapoints. Significance was calculated with the two-tailed unpaired Student's *t* test; ****P<0.0001. **(D)** Quantification of dystrophin transcripts in AdVP-edited muscle cells. RT-qPCR analysis of *DMD* expression on myotubes differentiated from human wild-type myoblasts initially transduced with AdVP.PE2$^{DMD.INS+1}$ or AdVP.PE3$^{DMD.DEL-2}$ at 400 TU per cell. Myotubes differentiated from mock-transduced myoblasts permitted measuring *DMD* mRNA steady-state levels. Significant differences between the indicated datasets were calculated with one-way ANOVA followed by Dunnett's multiple comparison tests; *0.01<*P*<0.05. Housekeeping *GAPDH* transcripts served as references for internal normalization of expression levels. **(E)** Assessing dystrophin knockout upon AdVP-based prime editing. Dystrophin western blotting was performed on myotubes differentiated from wild-type myoblasts transduced with the indicated *DMD* prime-editing AdVPs at 400 TU per cell (ten micrograms of total protein loaded per lane; 6% SDS-PAGE gel). Myotubes differentiated from mock-transduced myoblasts served as reference controls. Myogenic differentiation was controlled for by using an antibody directed to the late muscle-specific marker skeletal myosin heavy chain, and sample loading by applying antibodies recognizing vinculin and housekeeping GAPDH proteins.

By capitalizing on the cell cycle independency of adenovirus capsid-mediated delivery, we have previously found that PE2-based gene editing is, to some extent, hindered in non-cycling cells (27). To further investigate the role of the mitotic status of target cells on prime editing and, in particular, to compare PE2- versus PE3-based gene editing, AdVP.PE2$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.INS+1}$ were applied to cycling myoblasts and to their post-mitotic differentiated myotube counterparts (**Figure 4A**, left panel). Western blot analysis established similar amounts of prime editor proteins in myoblasts and myotubes transduced with either AdVP.PE2$^{DMD.INS+1}$ or AdVP.PE3$^{DMD.INS+1}$ (**Figure 4A**, left panel) yet, prime editing frequencies were significantly higher in myoblasts than in myotubes regardless of whether PE2 or PE3 setups were applied (**Figure 4A**, right panel). These data support the conclusion that cell cycling favours genomic DNA editions resulting from both 2- and 3-component prime editing systems.

In contrast to cells exposed to programable nucleases, the majority of cells subjected to prime editors retain unedited alleles intact offering the possibility for additional rounds of productive prime editing to enrich for precise genome editing events within target cell populations. Hence, we next sought to explore the simple transfection-independent and non-cytotoxic AdVP delivery process to test such prime-editing stacking approach based on sequential delivery of specific prime editing complexes (**Figure 4B**, left panel). Genotyping of *DMD* target alleles in human myoblasts transduced with AdVP.PE2$^{DMD.INS+1}$ by high-throughput NGS analysis did establish the gradual build-up of prime editing events in target cell populations subjected to three consecutive AdVP transduction rounds (**Figure 4B**, right panel). Hence, precise genetic modification of cell types amenable to *in vitro* culturing might profit from AdVP-assisted prime editing stacking especially in instances where chromosomal editing frequencies reach a plateau with a single delivery round due to refractory cellular or target sequence contexts.

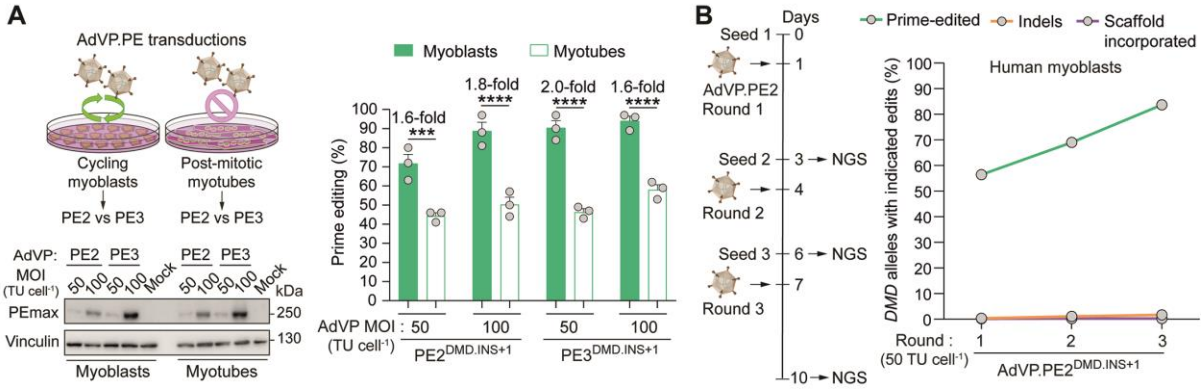


Figure 4. Assessing prime editing activities in cycling versus post-mitotic cells and upon sequential AdVP delivery. (A) Probing the impact of cell replication on PE2 versus PE3 systems. Experimental set-up and prime editor protein amounts in AdVP-transduced muscle cells (left panels). Western blot analysis of PEmax in muscle cells transduced before and after differentiation with AdVP.PE2$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.INS+1}$ at 50 or 100 transducing units (TU) per cell (thirty micrograms of total protein loaded per lane; 6% SDS-PAGE gel). Cas9- and vinculin-specific antibodies detected target PEmax and loading control proteins, respectively. Prime editing frequencies in mitotic versus post-mitotic muscle cells transduced with AdVP.PE2$^{DMD.INS+1}$ and AdVP.PE3$^{DMD.INS+1}$ were quantified by inference of CRISPR edits at 3 days post-transduction. Bars and error bars denote mean ± SEM, respectively, of three biological replicates (right panel). Significances between the indicated datasets were calculated with

two-way ANOVA followed by Tukey's multiple comparison tests; ****P<0.0001, ***0.0001<P<0.001. **(B)** Probing prime editing stacking upon sequential AdVP delivery. The buildup of prime editing events in myoblast populations after three transduction rounds with AdVP.PE2$^{DMD.INS+1}$ applied at 50 transducing units (TU) per cell was quantified by high-throughput NGS at the indicated timepoints.

DMD patients often succumb to the disease due to cardiac failure (28). The integration of advanced gene editing and human iPSC technologies offers the prospect for establishing relevant disease-in-a-dish systems to investigate DMD pathological processes and candidate therapeutic agents (38). In addition, iPSCs are promising substrates for DMD-directed cell therapies owing to their self-renewal and myogenic differentiation capabilities (49-51). Hence, to test AdVP-based prime editing in an iPSC disease-modelling context, iPSCs derived from a DMD patient with a DMD.Δ45-50 genotype were first triggered to differentiate into beating cardiomyocytes (**Supplementary Files** DOI: 10.6084/m9.figshare.24869136 https://figshare.com/s/848a70783590ab572bf0). Next, the differentiated cells were either exposed or not exposed to AdVP.PE3$^{DMD.INS+1}$ and, subsequently, were subjected to *DMD* editing and expression analyses (**Figure 5A**). The former analysis revealed a clear build-up of the programmed 1-bp insertion within *DMD* exon 51 (**Figure 5B**); and, consistently with this data, the latter analysis ascertained the induction of Becker-like dystrophin expression at the mRNA and protein levels (**Figure 5C** and **5D**, respectively).

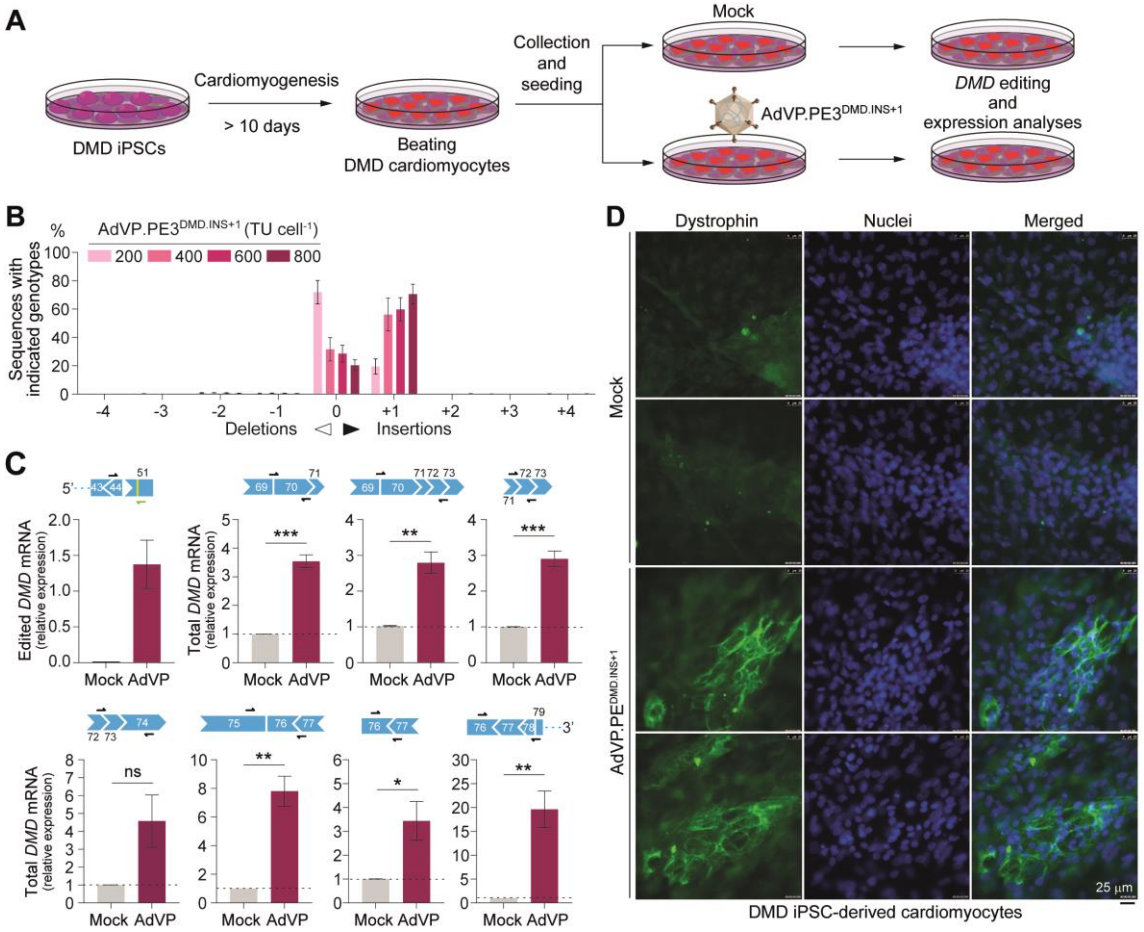


Figure 5. Testing AdVP-based prime editing in DMD iPSC-derived cardiomyocytes. (**A**) Illustration of the experimental setup. DMD iPSC-derived cardiomyocytes generated via a small-molecule differentiation protocol were transduced with prime-editing AdVP.PE3$^{DMD.INS+1}$ for endogenous *DMD* gene repair. As control, parallel cultures of differentiated cardiomyocytes were left untransduced. *DMD* editing and expression assays were performed at 4 days and 8 days post-transduction, respectively. (**B**) Quantification of prime editing. DMD iPSC-derived cardiomyocytes were transduced with AdVP.PE3$^{DMD.INS+1}$ at the indicatesd multiplicities of infection. Prime editing frequencies were determined through DNA sequencing genotyping assays at 4 days post-transduction. Bars and error bars correspond to, respectively, mean ± SEM from 3 biological replicates. (**C** and **D**) *DMD* expression analyses. RT-qPCR and fluorescence microscopy assays specific for *DMD* transcript and protein products, respectively, were done on cultures of cardiomyocytes differentiated from DMD iPSCs transduced with AdVP.PE3$^{DMD.INS+1}$ at 800 TU cell^{-1} at 8 days post-transduction. Parallel cultures of mock-transduced DMD iPSC-derived cardiomyocytes served to set *DMD* mRNA and protein baseline levels. Bars and error bars correspond to, respectively, mean ± SEM from three biological replicates. Significant differences between the indicated datasets were determined by two-tailed unpaired Student's t tests; ***0.0001 <P<0.001,**0.001<P<0.01, *0.01<P <0.05; P>0.05 was considered non-significant (ns).

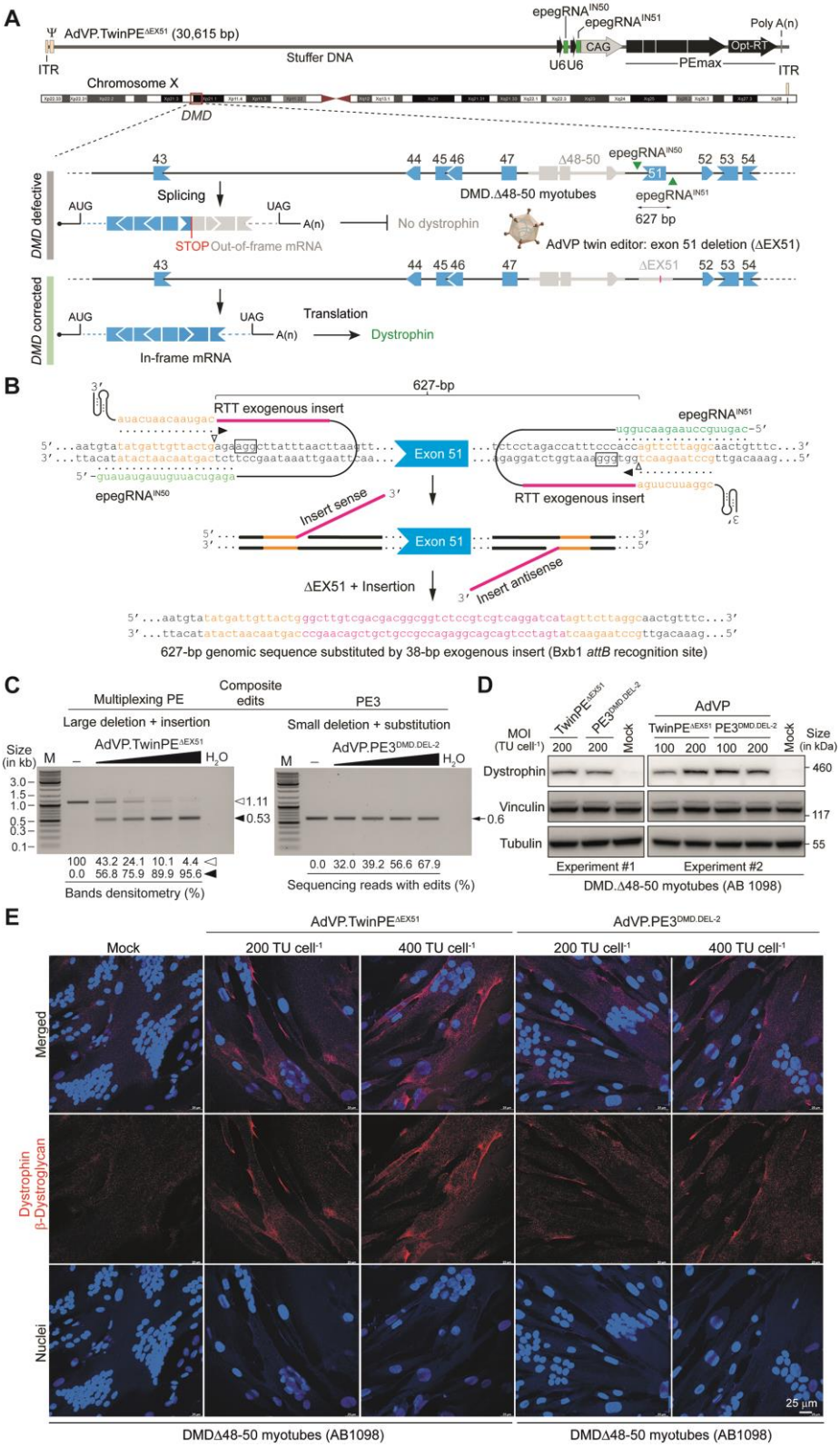


Figure 6. Gene correction through AdVP-based multiplexing prime editing in DMD defective myoblasts. (**A**) Genome structure of AdVP assembled for *DMD* gene correction using multiplexing prime editing complexes. ITR and Ψ, adenovirus type-5 cis-acting inverted terminal repeats and packaging signal, respectively. The hybrid CAG promoter drives PEmax synthesis whilst human U6 promoters drive the expression of a epegRNA pair (i.e., epegRNA^{IN50} and epegRNA^{IN51}) for *DMD* reading frame repair in muscle cells amenable to exon 51 excision (e.g., DMD.Δ48-50). (**B**) Schematics of *DMD* exon 51 excision through twin prime editing. Spacer, primer binding site (PBS) and reverse transcriptase template (RTT) sequences of epegRNA^{IN50} and epegRNA^{IN51} are highlighted in green, orange and magenta, respectively. The latter sequence encodes exogenous genetic information in the form of the serine recombinase Bxb1 *attB* recognition site. Protospacer adjacent motifs (NGG) are boxed, and nicking positions are marked by open arrowheads. Twin prime editors engage offset protospacer sequences on opposite DNA strands generating nicks that lead to the hybridization of the released single-stranded DNA strands to each PBS. The resulting free 3' hydroxyl groups

prime the synthesis of 3' DNA flaps over RTT sequences by the reverse transcriptases. After the annealing of 3' and 5' DNA flaps containing edited and original DNA sequences (not shown), respectively, removal of the 5' flaps followed by ligation of the 3' flaps to the respective DNA excising nicks yields the intended gene-editing product, i.e., replacement of genomic DNA encompassing *DMD* exon 51 by the Bxb1 *attB* recognition site. (**C**) Testing AdVP delivery of functional prime-editing multiplexes. Human myoblasts with a Δ48-50 genotype (AB1098) were transduced with AdVP.TwinPE$^{\Delta EX51}$ or AdVP.PE3$^{DMD.DEL-2}$ at 50, 100, 200 and 400 TU cell^{-1}. Twin PE- and PE3-derived prime edits were traced at 3 days post-transduction by DNA densitometry and sequencing of target DNA amplicons, respectively. (**D**) Dystrophin detection in DMD muscle cells corrected via AdVP delivery of prime-editing multiplexes. Western blotting was performed on myotubes differentiated from DMD.Δ48-50 myoblasts previously transduced with the indicated *DMD* prime-editing AdVPs (sixty micrograms of total protein loaded per lane; 6% SDS-PAGE gel). Detection of vinculin and tubulin provided for independent protein loading controls. (**E**) Dystrophin-β-dystroglycan interaction analysis in DMD muscle cells after AdVP transfer of single and dual prime-editing complexes. Detection of endogenous dystrophin-β-dystroglycan interactions by proximity ligation assays on myotubes differentiated from DMD.Δ48–50 myoblasts transduced with the indicated *DMD* prime-editing AdVPs (red foci). Parallel cultures of untreated DMD.Δ48–50 myotubes (Mock) served negative controls. Nuclei were labelled by DAPI staining.

The recent development of multiplexing prime editing strategies based on the delivery of prime editors and dual pegRNAs is contributing to further expand the scope of DSB-independent genome editing (2-7). In particular, via targeting offset target sites on opposite DNA strands and locally reverse-transcribing complementary DNA sequences, pairs of prime editing complexes are capable of yielding genomic insertions, deletions and/or substitutions whose sizes are substantially larger than those enabled via the use of PE2 and PE3 components (2-7). To investigate the feasibility and utility of AdVP-based multiplexing prime editing, the vector AdVP.TwinPE$^{\Delta EX51}$ was assembled. This vector encodes PEmax and dual epegRNAs whose "twin" arrangement (2) is designed for DMD exon 51 deletion and concomitant insertion of a recombinase recognition site (**Figure 6A** and **6B**). Of notice, most DMD-causing mutations cluster inside the exon 45-55 region (major *DMD* mutational hotspot) with the majority of these, underlying circa 13% of all DMD cases, being amenable to repair through exon 51 skipping or deletion (28) (**Figure 6B**).

Similarly to AdVPs encoding PE2 and PE3 components, AdVP.TwinPE$^{\Delta EX51}$ packaged structurally intact vector genomes (**Supplementary Figure 16**) and was produced to an high titre (i.e., 1.81×10^{10} TU ml^{-1}). Crucially, transduction experiments testing AdVP.TwinPE$^{\Delta EX51}$ next to AdVP.PE3$^{DMD.DEL-2}$ in human myoblasts with a DMD.Δ48-50 genotype, established the functionality of AdVP-delivered prime editing multiplexes via the detection of a dose-dependent accumulation of genomic edits encompassing the intended DMD exon 51 deletion (**Figure 6C** and **Supplementary Figure 17**). In fact, differentiation of DMD.Δ48-50 muscle progenitors that had been prime-edited through AdVP.TwinPE$^{\Delta EX51}$ and AdVP.PE3$^{DMD.DEL-2}$ both readily led to the detection of Becker-like dystrophins (**Figure 6D**) as well as to the assembly of protein complexes connecting these dystrophin molecules to its DGC partner β-dystroglycan (**Figure 6E**).

Taken together, these experiments support the suitability and versatility of AdVP-based prime editing for disease modelling as well as for precise gene knockout or correction in human stem/progenitor cells and their differentiated progenies.

DISCUSSION

Cell and gene therapies for DMD are under intense investigation and include the transplantation of ex vivo corrected myogenic cells and the *in vivo* delivery of RNA-guided nucleases, respectively (28,49-52). Clearly, each of these modalities have their own sets of pros and cons (49). For example, although *ex vivo* approaches offer a controlled gene repair setting and minimize immune responses to vector and gene-editing tool components, they currently present notable bottlenecks, e.g., limited cell survival and tissue engraftment (49-51). Hence, *in vivo* DMD-directed therapeutic modalities such as those based on co-administering dual AAVs encoding Cas9 nucleases and cognate gRNAs, are also being actively investigated (52). Despite the detection of immune responses against capsid and nuclease components in adult immunocompetent animals, collectively, these reports demonstrate that AAV-based *DMD* gene repair can improve striated muscle function. A potentially insidious outcome identified is, however, the prevalent integration of AAV vector DNA at site-specific DSBs, including at *Dmd* exons 51 and 53 in muscle tissues (53,54). These data stress the need to expand candidate genetic therapies to DSB-free gene editing systems as those based on base editors and prime editors. Prototypic base editors comprise a regular gRNA and a Cas9^{D10A} nickase linked to a cytidine or adenine deaminase that, upon target nucleotide deamination and subsequent DNA repair or replication, yield C→T and A→G substitutions, respectively (55-57). Owing to their dependency on regular gRNAs, it is easier designing

and identifying robust base editors than prime editors, provided that a PAM exists for placing a target nucleotide within the base editor's activity window. On the contrary, besides being more prone to off-target genomic modifications than prime editors and limited to installing single base-pair substitutions, base editors create bystander edits if non-target nucleotide(s) locate within their activity windows. Hence, the powerful and, to some extent, complementary attributes of base editing and prime editing technologies is spurring their research and development. In this context, delivery systems based on dual AAV strategies comprising two AAV vectors each encoding split portions of prime editors or base editors are being actively pursued. In cells co-transduced with split AAV vectors, prime editing or base editing ensues upon *in situ* assembly of complete proteins via intein trans-splicing dependent and independent processes (11-20). A recently optimized dual AAV prime editing system yielded up to 11% of precise gene edition in murine hearts (58). Moreover, dual AAV base editing systems were shown to, either via targeted splice site motif disruption or point mutation correction, yield *Dmd* reading frame repair and ensuing dystrophin expression in striated muscles of dystrophic mice (59,60).

Notwithstanding the amassing of important proof-of-concepts for disease modelling and gene correction, dual AAV designs are complex and require that co-transductions lead to effective and proper assembly of independent gene-editing tool parts. Recently, adenovectors deleted in early viral genes and encoding a shortened prime editor lacking the dispensable RNaseH domain, were shown to be superior to dual AAV vectors for prime editing in mouse livers (14). However, high immunogenicity *in vivo* and cytotoxicity *in vitro* is often associated with these first-generation adenoviral vectors due to their high viral gene content (24-26). Hence, there is also a pressing need to expand the range of prime editing delivery options, especially those that like AAV lack viral genes but that instead of AAV have large cargo capacities.

Towards this goal, in this study, we have established the feasibility of deploying fully viral gene-deleted AdVPs for efficient *DMD* prime editing in cell types whose myogenic capacity has supported their investigation as candidate cell therapy substrates (i.e., myoblasts, mesenchymal stem cells and iPSCs) (50-51). Indeed, combined all-in-one AdVP transfer of optimized prime-editing components in the form of PEmax (8), DMD-targeting epegRNAs (9) and auxiliary gRNAs with an improved scaffold (61), or optimized dual prime editing complexes, resulted in the robust accumulation of *DMD* edits in the form of precise small insertions or deletions or whole exon excisions. Importantly, DMD myoblasts subjected to AdVP delivery of optimized prime-editing complexes retained their myogenic differentiation capacity resulting in selection-free detection of Becker-like dystrophin molecules capable of physically associating with β-dystroglycan, a key component of the DGC (46). Moreover, gene knockout and gene repair experiments in skeletal muscle and iPSC-derived cardiac cells, respectively, have further supported AdVP-based prime editing for establishing human disease-in-a-dish models that can be directed for studying pathological processes or screening therapeutic candidates.

Despite our finding that PE2- and PE3-based gene editing is somewhat less active in post-mitotic than in cycling muscle cells, the capacity of AdVPs to efficiently transduce cells independently of their mitotic status warrants their future testing in animal models, including in humanized dystrophin-defective mice in which human DMD gene-tailored tools can be directly tested *in vivo*. Finally, we demonstrate that combining facile and non-cytotoxic AdVP transduction with non-mutagenic prime editing, permits the selective stacking of precise genome editing events in target cell populations via reiterated delivery of prime editing complexes. Such protocols might be beneficial in instances where prime editing reaches a single-dose plateau or is suboptimal due to refractory cellular or target site contexts.

Precision genome editing is increasingly underpinned by large and multicomponent tools whose testing and application using common delivery agents such as AAV is rendered complex or ineffective. Moreover, in a recent study from our laboratory investigating AdVP transfer of forced CRISPR-Cas9 heterodimers, it is demonstrated that the efficiency and accuracy of multiplexing genome editing can profit from integrated as opposed to separated delivery of the attendant reagents (62). Presumably, these improved outcomes result from the increased likelihood that integrated delivery leads to a more balanced assembly and synchronous action of otherwise individually acting CRISPR-Cas9 complexes. Hence, it is possible that other advanced multiplexing genome editing approaches will equally profit from combined all-in-one delivery systems. Amongst these systems are those based on prime editors and dual pegRNAs that, via targeting bipartite target sites on opposite strands and reverse-transcribing

complementary DNA sequences, permit deleting or replacing specific genomic tracts (2-7). Indeed, in this study, we demonstrate the value of AdVP-assisted multiplexing prime editing based on all-in-one delivery of full-length PEmax and dual epegRNAs designed for *DMD* reading frame repair through targeted exon deletion.

Finally, as aforementioned, experimental data reported elsewhere and here links prime editing performance to cell type- or cell stage-specific determinants, namely, complement of DNA repair factors (8,22), cell-cycle activity (**Figure 4**) (22,27) and/or chromatin context (45). The tropism adaptability and cell cycle independency of AdVP systems might thus facilitate probing wanted and unwanted effects of specific prime editing reagents directly on the (epi)genomes of different human cell types at mature, progenitor or undifferentiated stages.

In conclusion, combined delivery of full-length conventional and dual prime editing components in single AdVPs yields efficient and precise modification of *DMD* alleles in stem/progenitor cells with myogenic capacity. Generically, AdVPs serve as a robust and versatile platform for investigating advanced prime editing principles in difficult-to-transfect cell types independently of the size and numbers of the attendant reagents. As a corollary, AdVP-assisted prime editing warrants further research and testing, including for the modelling and repairing of genetic defects underlying human disorders in *ex vivo* and *in vivo* settings.

DATA AVAILABILITY

The data supporting this study are included in the article and accompanying supplementary files. The raw deep-sequencing library reads are deposited at the NCBI Sequence Read Archive (SRA) database under BioProject ID number: PRJNA957967. The movies illustrating light microscopy fields of beating DMD iPSC-derived cardiomyocytes are available in Figshare at https://doi.org/10.6084/m9.figshare.24869136.

ACKNOWLEDGMENTS

The authors thank Vincent Mouly (Institute of Myology, Sorbonne University, Paris, France) for providing the human myoblasts with wild-type and DMD defective genotypes and Chuannan Fan (LUMC, Department of Cell and Chemical Biology) for performing western blots. The authors are also thankful to Antoine A.F. de Vries (LUMC, Department of Cardiology) for providing a construct encoding the GpNLuc reporter. Authors of this study are members of the European Reference Network – Neuromuscular diseases (ERN EURO-NMD).

FUNDING

This research was supported by the Prinses Beatrix Spierfonds (grant number: W.OR21-01). This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska Curie Actions grant agreement no. 101072427 (GetRadi – Gene Therapy of Rare Diseases). Qian Wang held a Ph.D. fellowship from the China Scholarship Council–Leiden University joint scholarship programme.
Conflict of interest statement. None declared.

SUPPLEMENTARY DATA

Supplementary Data are available online at doi: 10.1093/nar/gkae057.

REFERENCES

1. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. et al. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 576, 149-157.
2. Anzalone, A.V., Gao, X.D., Podracky, C.J., Nelson, A.T., Koblan, L.W., Raguram, A., Levy, J.M., Mercer, J.A.M. and Liu, D.R. (2022) Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat Biotechnol, 40, 731-740.
3. Choi, J., Chen, W., Suiter, C.C., Lee, C., Chardon, F.M., Yang, W., Leith, A., Daza, R.M., Martin, B. and Shendure, J. (2022) Precise genomic deletions using paired prime editing. Nat Biotechnol, 40, 218-226.

4. Jiang, T., Zhang, X.O., Weng, Z. and Xue, W. (2022) Deletion and replacement of long genomic sequences using prime editing. Nat Biotechnol, 40, 227-234.
5. Tao, R., Wang, Y., Jiao, Y., Hu, Y., Li, L., Jiang, L., Zhou, L., Qu, J., Chen, Q. and Yao, S. (2022) Bi-PE: bi-directional priming improves CRISPR/Cas9 prime editing in mammalian cells. Nucleic Acids Res, 50, 6423-6434.
6. Wang, J., He, Z., Wang, G., Zhang, R., Duan, J., Gao, P., Lei, X., Qiu, H., Zhang, C., Zhang, Y. et al. (2022) Efficient targeted insertion of large DNA fragments without DNA donors. Nat Methods, 19, 331-340.
7. Zhuang, Y., Liu, J., Wu, H., Zhu, Q., Yan, Y., Meng, H., Chen, P.R. and Yi, C. (2022) Increasing the efficiency and precision of prime editing with guide RNA pairs. Nat Chem Biol, 18, 29-37.
8. Chen, P.J., Hussmann, J.A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.F., Chen, C., Nelson, J.W., Newby, G.A., Sahin, M. et al. (2021) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell, 184, 5635-5652.
9. Nelson, J.W., Randolph, P.B., Shen, S.P., Everette, K.A., Chen, P.J., Anzalone, A.V., An, M., Newby, G.A., Chen, J.C., Hsu, A. et al. (2022) Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol, 40, 402-410.
10. Chen, P.J. and Liu, D.R. (2023) Prime editing for precise and highly versatile genome manipulation. Nat Rev Genet, 24, 161-177.
11. Liu, P., Liang, S.Q., Zheng, C., Mintzer, E., Zhao, Y.G., Ponnienselvan, K., Mir, A., Sontheimer, E.J., Gao, G., Flotte, T.R. et al. (2021) Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. Nat Commun, 12, 2121.
12. Jang, H., Jo, D.H., Cho, C.S., Shin, J.H., Seo, J.H., Yu, G., Gopalappa, R., Kim, D., Cho, S.R., Kim, J.H. et al. (2022) Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. Nat Biomed Eng, 6, 181-194.
13. Zhi, S., Chen, Y., Wu, G., Wen, J., Wu, J., Liu, Q., Li, Y., Kang, R., Hu, S., Wang, J. et al. (2022) Dual-AAV delivering split prime editor system for in vivo genome editing. Mol Ther, 30, 283-294.
14. Bock, D., Rothgangl, T., Villiger, L., Schmidheini, L., Matsushita, M., Mathis, N., Ioannidi, E., Rimann, N., Grisch-Chan, H.M., Kreutzer, S. et al. (2022) In vivo prime editing of a metabolic liver disease in mice. Sci Transl Med, 14, eabl9238.
15. Gao, Z., Ravendran, S., Mikkelsen, N.S., Haldrup, J., Cai, H., Ding, X., Paludan, S.R., Thomsen, M.K., Mikkelsen, J.G. and Bak, R.O. (2022) A truncated reverse transcriptase enhances prime editing by split AAV vectors. Mol Ther, 30, 2942-2951.
16. She, K., Liu, Y., Zhao, Q., Jin, X., Yang, Y., Su, J., Li, R., Song, L., Xiao, J., Yao, S. et al. (2023) Dual-AAV split prime editor corrects the mutation and phenotype in mice with inherited retinal degeneration. Signal Transduct Target Ther, 8, 57.
17. Qin, H., Zhang, W., Zhang, S., Feng, Y., Xu, W., Qi, J., Zhang, Q., Xu, C., Liu, S., Zhang, J. et al. (2023) Vision rescue via unconstrained in vivo prime editing in degenerating neural retinas. J Exp Med, 220, e20220776.
18. Liu, B., Dong, X., Cheng, H., Zheng, C., Chen, Z., Rodriguez, T.C., Liang, S.Q., Xue, W. and Sontheimer, E.J. (2022) A split prime editor with untethered reverse transcriptase and circular RNA template. Nat Biotechnol, 40, 1388-1393.
19. Zheng, C., Liang, S.Q., Liu, B., Liu, P., Kwan, S.Y., Wolfe, S.A. and Xue, W. (2022) A flexible split prime editor using truncated reverse transcriptase improves dual-AAV delivery in mouse liver. Mol Ther, 30, 1343-1351.
20. Grunewald, J., Miller, B.R., Szalay, R.N., Cabeceiras, P.K., Woodilla, C.J., Holtz, E.J.B., Petri, K. and Joung, J.K. (2023) Engineered CRISPR prime editors with compact, untethered reverse transcriptases. Nat Biotechnol, 41, 337-343.
21. Levesque, S., Mayorga, D., Fiset, J.P., Goupil, C., Duringer, A., Loiselle, A., Bouchard, E., Agudelo, D. and Doyon, Y. (2022) Marker-free co-selection for successive rounds of prime editing in human cells. Nat Commun, 13, 5909.
22. Schene, I.F., Joore, I.P., Baijens, J.H.L., Stevelink, R., Kok, G., Shehata, S., Ilcken, E.F., Nieuwenhuis, E.C.M., Bolhuis, D.P., van Rees, R.C.M. et al. (2022) Mutation-specific reporter for optimization and enrichment of prime editing. Nat Commun, 13, 1028.
23. Simon, D.A., Talas, A., Kulcsar, P.I., Biczok, Z., Krausz, S.L., Varady, G. and Welker, E. (2022) PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. Elife, 11, e69504.

24. Gonçalves, M.A. and de Vries, A.A. (2006) Adenovirus: from foe to friend. Rev Med Virol, 16, 167-186.
25. Tasca, F., Wang, Q. and Gonçalves, M. (2020) Adenoviral vectors meet gene editing: a rising partnership for the genomic engineering of human stem cells and their progeny. Cells, 9, 953.
26. Ricobaraza, A., Gonzalez-Aparicio, M., Mora-Jimenez, L., Lumbreras, S. and Hernandez-Alcoceba, R. (2020) High-Capacity Adenoviral Vectors: Expanding the Scope of Gene Therapy. Int J Mol Sci, 21, 3643.
27. Wang, Q., Liu, J., Janssen, J.M., Tasca, F., Mei, H. and Gonçalves, M.A.. (2021) Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. Nucleic Acids Res, 49, 11986-12001.
28. Duan, D., Goemans, N., Takeda, S., Mercuri, E. and Aartsma-Rus, A. (2021) Duchenne muscular dystrophy. Nat Rev Dis Primers, 7, 13.
29. Tasca, F., Brescia, M., Wang, Q., Liu, J., Janssen, J.M., Szuhai, K. and Gonçalves M.A.(2022) Large-scale genome editing based on high-capacity adenovectors and CRISPR-Cas9 nucleases rescues full-length dystrophin synthesis in DMD muscle cells. Nucleic Acids Res., 50, 7761-7782.
30. Gonçalves, M.A., van der Velde, I., Janssen, J.M., Maassen, B.T., Heemskerk, E.H., Opstelten, D.J., Knaan-Shanzer, S., Valerio, D. and de Vries, A.A. (2002) Efficient generation and amplification of high-capacity adeno-associated virus/adenovirus hybrid vectors. J Virol, 76, 10734-10744.
31. Mamchaoui, K., Trollet, C., Bigot, A., Negroni, E., Chaouch, S., Wolff, A., Kandalla, P.K., Marie, S., Di Santo, J., St Guily, J.L. et al. (2011) Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. Skelet Muscle, 1, 34.
32. Thorley, M., Duguez, S., Mazza, E.M.C., Valsoni, S., Bigot, A., Mamchaoui, K., Harmon, B., Voit, T., Mouly, V. and Duddy, W. (2016) Skeletal muscle characteristics are preserved in hTERT/cdk4 human myogenic cell lines. Skelet Muscle, 6, 43.
33. Gonçalves, M.A., Swildens, J., Holkers, M., Narain, A., van Nierop, G.P., van de Watering, M.J., Knaan-Shanzer, S. and de Vries, A.A. (2008) Genetic complementation of human muscle cells via directed stem cell fusion. Mol Ther, 16, 741-748.
34. Wang, Q., Liu, J., Janssen, J.M., Le Bouteiller, M., Frock, R.L. and Gonçalves, M. (2021) Precise and broad scope genome editing based on high-specificity Cas9 nickases. Nucleic Acids Res, 49, 1173-1198.
35. Fallaux, F.J., Bout, A., van der Velde, I., van den Wollenberg, D.J., Hehir, K.M., Keegan, J., Auger, C., Cramer, S.J., van Ormondt, H., van der Eb, A.J. et al. (1998) New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. Hum Gene Ther, 9, 1909-1917.
36. Schaub, F.X., Reza, M.S., Flaveny, C.A., Li, W., Musicant, A.M., Hoxha, S., Guo, M., Cleveland, J.L. and Amelio, A.L. (2015) Fluorophore-NanoLuc BRET Reporters Enable Sensitive In Vivo Optical Imaging and Flow Cytometry for Monitoring Tumorigenesis. Cancer Res., 75, 5023-33.
37. Janssen, J.M., Liu, J., Skokan, J., Gonçalves, M.A. and de Vries, A.A. (2013) Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. J Gene Med, 15, 1-11.
38. Campostrini, G., Meraviglia, V., Giacomelli, E., van Helden, R.W.J., Yiangou, L., Davis, R.P., Bellin, M., Orlova, V.V. and Mummery, C.L. (2021) Generation, functional analysis and applications of isogenic three-dimensional self-aggregating cardiac microtissues from human pluripotent stem cells. Nat. Protoc.,16, 2213-2256.
39. Conant, D., Hsiau, T., Rossi, N., Oki, J., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K. et al. (2022) Inference of CRISPR Edits from Sanger Trace Data. CRISPR J, 5, 123-130.
40. Brinkman, E.K., Chen, T., Amendola, M. and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res, 42, e168.
41. Concordet, J.P and Haeussler, M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res., 46, W242-W245.
42. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E. et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol, 37, 224-226.
43. Knaan-Shanzer, S., van de Watering, M.J., van der Velde, I., Gonçalves, M.A., Valerio, D. and de Vries, A.A. (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.
44. Gonçalves, M.A., Holkers, M., Cudre-Mauroux, C., van Nierop, G.P., Knaan-Shanzer, S., van der Velde, I., Valerio, D. and de Vries, A.A. (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.
45. Park, S.J., Jeong, T.Y., Shin, S.K., Yoon, D.E., Lim, S.Y., Kim, S.P., Choi, J., Lee, H., Hong, J.I., Ahn, J. et al. (2021) Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. Genome Biol, 22, 170.
46. Matsumura, K. and Campbell, K.P. (1993) Deficiency of dystrophin-associated proteins: a common mechanism leading to muscle cell necrosis in severe childhood muscular dystrophies. Neuromuscul Disord, 3, 109-118.
47. Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc Natl Acad Sci U S A, 98, 8241-8246.
48. Chen, X., Tasca, F., Wang, Q., Liu, J., Janssen, J.M., Brescia, M.D., Bellin, M., Szuhai, K., Kenrick, J., Frock, R.L. et al. (2020) Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking. Nucleic Acids Res, 48, 974-995.
49. Maggio, I., Chen, X. and Gonçalves, M.A. (2016) The emerging role of viral vectors as vehicles for DMD gene editing. Genome Med, 8, 59.
50. Biressi, S., Filareto, A. and Rando, T.A. (2020) Stem cell therapy for muscular dystrophies. J. Clin. Invest., 130, 5652-5664.
51. Boyer, O., Butler-Browne, G., Chinoy, H., Cossu, G., Galli, F., Lilleker, J.B., Magli, A., Mouly, V., Perlingeiro, R.C.R., Previtali, S.C. et al. (2021) Myogenic Cell Transplantation in Genetic and Acquired Diseases of Skeletal Muscle. Front. Genet., 12, 702547.
52. Chemello, F., Bassel-Duby, R. and Olson, E.N. (2020) Correction of muscular dystrophies by CRISPR gene editing. J Clin Invest, 130, 2766-2776.
53. Nelson, C.E., Wu, Y., Gemberling, M.P., Oliver, M.L., Waller, M.A., Bohning, J.D., Robinson-Hamm, J.N., Bulaklak, K., Castellanos Rivera, R.M., Collier, J.H. et al. (2019) Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. Nat Med, 25, 427-432.
54. Hanlon, K.S., Kleinstiver, B.P., Garcia, S.P., Zaborowski, M.P., Volak, A., Spirig, S.E., Muller, A., Sousa, A.A., Tsai, S.Q., Bengtsson, N.E. et al. (2019) High levels of AAV vector integration into CRISPR-induced DNA breaks. Nat Commun, 10, 4439.
55. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature, 533, 420-424.
56. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K,Y. et al. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science, 353, aaf8729.
57. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I. and Liu, D.R. (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature, 551, 464-471.
58. Davis, J.R., Banskota, S., Levy, J.M., Newby, G.A., Wang, X., Anzalone, A.V., Nelson, A.T., Chen, P.J., Hennes, A.D., An, M. et al. (2023) Efficient prime editing in mouse brain, liver and heart with dual AAVs. Nat. Biotechnol. doi: 10.1038/s41587-023-01758-z. Online ahead of print.
59. Chemello, F., Chai, A.C., Li, H., Rodriguez-Caycedo, C., Sanchez-Ortiz, E., Atmanli, A., Mireault, A.A., Liu, N., Bassel-Duby, R. and Olson, E.N. (2021) Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. Sci. Adv., 7, eabg4910.
60. Xu, L., Zhang, C., Li, H., Wang, P., Gao, Y., Mokadam, N.A., Ma, J., Arnold, W.D. and Han, R. (2021) Efficient precise in vivo base editing in adult dystrophic mice. Nat. Commun., 12, 3719.
61. Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P. and Wu, H. (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol, 16, 280.
62. Tasca, F., Brescia, M., Liu, J., Janssen, J.M., Mamchaoui, K. and Gonçalves, M. (2023) High-capacity adenovector delivery of forced CRISPR-Cas9 heterodimers fosters precise chromosomal deletions in human cells. Mol Ther Nucleic Acids, 31, 746-762.

Chapter 7

The chromatin context differently impacts prime editors and base editors and further controls the fidelity and purity of base editing

Qian Wang[1,†], Hidde Zittersteijn[1,†], Xiaoling Wang[1], Josephine M. Janssen[1], Jin Liu[1], Leon Mei[2], Rob C. Hoeben[1] and Manuel A.F.V. Gonçalves[1]

[1]Department of Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands;
[2]Department of Biomedical Data Sciences, Sequencing Analysis Support Core, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands
[†]These authors contributed equally

Under revision at Molecular Cell

ABSTRACT

Base editors and prime editors allow changing specific nucleotide sequences within the vast genomes of eukaryotic cells requiring neither mutagenic double-stranded DNA breaks nor exogenous donor DNA substrates. However, the performance of base editors vis-à-vis prime editors at alternate chromatin states is ill-defined. Moreover, the role of the chromatin environment of target sequences and its underlying factors on DNA editing product fidelity and purity is equally unknown. Here, using cellular systems that permit assessing the efficiency and fidelity of gene-editing tools at isogenic target sequences controlled by specific epigenetic factors, we report that heterochromatin impinged by the KRAB/KAP-1/HP1 axis alone or together with the DNA methyltransferases DNMT3A and DNMT3L, mostly hinders prime editors over base editors with the extended portions of prime-editing guide RNAs contributing to this outcome. Indeed, the performance of base editors at heterochromatin ranges in a target site-dependent manner from lower to, often, significantly higher than that observed at euchromatin. Additionally, the extent and types of byproducts accumulated after base editing is also contingent upon the epigenetic context of target sequences. Our findings have direct implications for the optimal assessment of these powerful genomic engineering tools and might guide their selection, further development and application.

INTRODUCTION

Genome editing based on CRISPR-associated (Cas) nucleases and sequence-customizable single guide RNAs (gRNAs) has become a powerful approach for introducing specific genetic changes (edits) in living cells[1]. However, in addition to the intended edits, repair of double-stranded DNA breaks (DSBs) by error-prone recombination processes frequently yields unwanted byproducts in the form of uncontrolled insertions and deletions (indels)[2-5], genome-wide translocations[6-10] and gross on-target chromosomal rearrangements[6,11]. Recent studies have also uncovered loss-of-heterozygosity, chromosome fragmentation followed by haphazard DNA reassembly (chromothripsis)**,** and whole chromosome losses (aneuploidy) upon target DSB formation[12-14]. Thus, although emerging high-specificity nucleases present reduced off-target activities[1,15,16], they are inherently incapable of eliminating the unintended and poorly controlled effects resulting from on-target DSBs. Therefore, increasing research is directed to substituting programmable nucleases by DSB-free genome editing systems, such as those based on Cas9 nickases as such[17-20], or on these nickases fused to DNA modifying effector domains that form base editors[21-24] and, more recently, prime editors[25].

Base editing complexes comprise a conventional gRNA and a Cas9^{D10A} nickase (**Supplementary Figure S1**) fused to cytidine or adenine deaminases[21-24]. Deaminated nucleotides generated *in situ* by cytidine base editors (CBEs) and adenine base editors (ABEs) are processed through DNA repair mechanisms ultimately yielding C•G-to-T•A (C→T) and A•T-to-G•C (A→G) transitions, respectively (**Supplementary Figure S1**). These base-pair substitutions take place prevalently within a so-called "editing window" whose length and location in the gRNA target sequence (protospacer) depends on the particular base editor architecture[24].

Prime editing complexes consist of an extended gRNA, named pegRNA, and a Cas9^{H840A} nickase fused to an engineered reverse transcriptase (RT) (**Supplementary Figure S2**). The pegRNA is formed by a gRNA covalently linked to RT template and primer binding site (PBS) sequences. Targeted nicking by Cas9^{H840A} releases a DNA flap that, upon annealing to the PBS, primes reverse transcription over the RT template that encodes the edit-of-interest. Through a series of cellular processing steps, the resulting DNA copy becomes ultimately incorporated at the genomic target site (**Supplementary Figure S2**). Although detailed investigation on the late-stage processing steps is required, DNA mismatch repair factors and cellular replication were recently shown to be determinants of prime editing[26-28]. Prime editing has two generic modalities, namely, PE2 and PE3. The former system relies exclusively on PE2:pegRNA complexes; the latter depends on the concerted action of PE2:pegRNA and PE2:gRNA complexes (**Supplementary Figure S2**). The PE3 system has enhanced activity, although the nicking of both DNA strands by PE3 components can foster indel byproduct accumulation[25].

Base editors are restricted to installing specific base-pair substitutions, whilst prime editors install well-defined insertions and deletions in addition to all 12 base-pair substitutions and combinations thereof[25]. Moreover, CBE and ABE deaminase effectors often do not discriminate target from nearby non-target nucleotides and can install unintended substitutions leading to reduced product purity[24]. Conversely,

base editors yield low indel byproducts and are normally more robust than prime editors at randomly selected target sequences. Hence, base editors and prime editors present a rather complementary set of attributes in terms of their editing versatility, robustness, and fidelity. It is, therefore, essential to identify the parameters underlying the individual and relative performances of base editors and primer editors to guide their further development and selection to specific contexts and goals.

Research from our laboratory and that from others has demonstrated that the activity of different types of nucleases, including CRISPR-derived nucleases, are significantly hindered by heterochromatic states in living cells[29-32]. However, cause-effect associations between alternate chromatin conformations and the activity and fidelity of DSB-free genome editing platforms remain to be characterized and thoroughly assessed. Hence, in this work, we sought to address these knowledge gaps by implementing complementary loss-of-function and gain-of-function cellular systems in which isogenic target sequences acquire specific euchromatic and heterochromatic statuses through the controlled recruitment of endogenous epigenetic remodelling complexes. We report that primer editing is frequently hindered at Krüppel-associated box (KRAB)-impinged facultative heterochromatin as well as at heritable heterochromatin created by the concerted action of KRAB-recruited remodelling complexes and DNA methyltransferases (i.e., DNMT3A and DNMT3L). Moreover, we found that the underperformance of prime editors at heterochromatic sequences is contributed by their pegRNA component. In contrast, for most target sequences tested, base editing activities were similar at euchromatin and heterochromatin or, often, were even higher at the latter closed chromatin conformation. Finally, our experiments reveal that not only base editing activity as such, but also the proportions between different types of base-editing byproducts are dependent on the epigenetic status of target sequences.

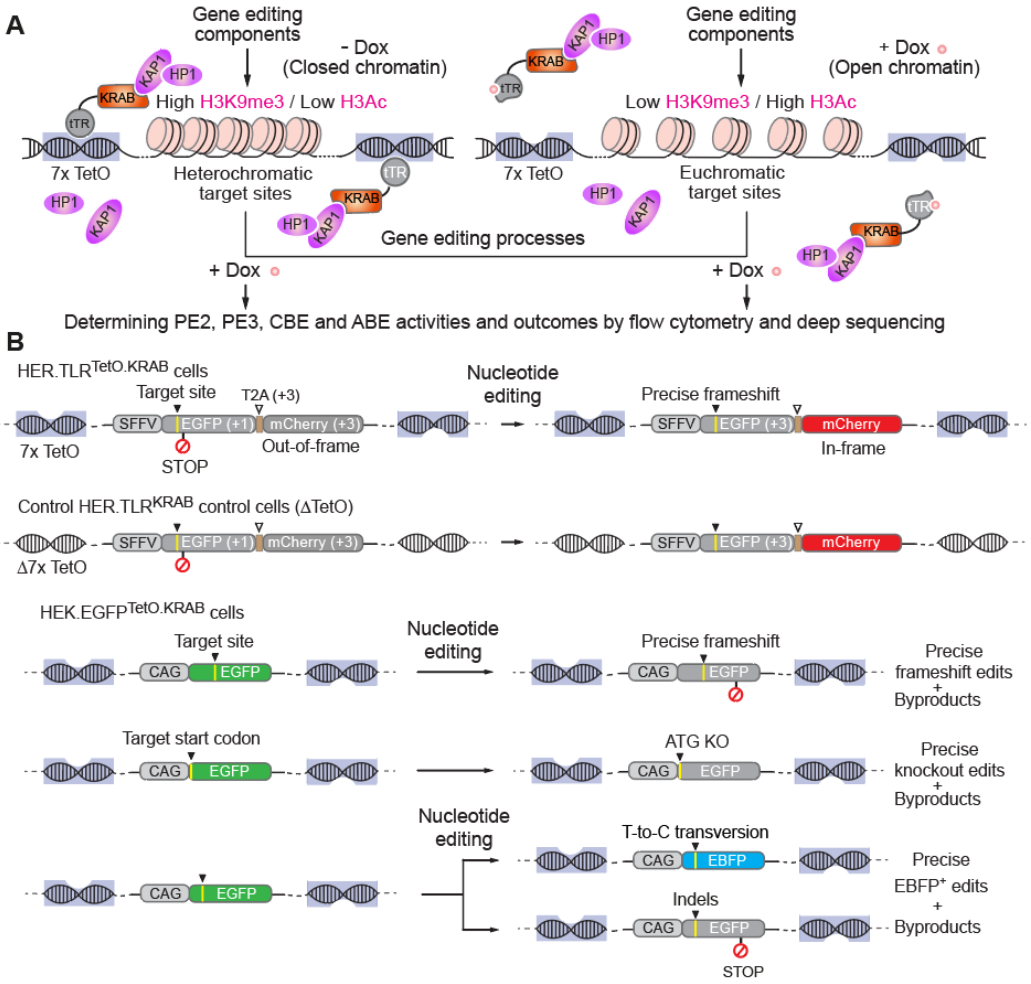


Figure 1. Experimental systems for tracing gene-editing activities and outcomes at isogenic target sequences at different chromatin conformations. (A) General experimental set-ups. Human reporter cells HER.TLR$^{TetO.KRAB}$ and HEK.EGFP$^{TetO.KRAB}$ cultured without or with doxycycline (Dox), are exposed to DSB-free gene editing tools in the form of PE2:pegRNA prime editing

complexes alone (PE2 system) or together with an auxiliary gRNA (PE3 system) or to CBE:gRNA or ABE:gRNA base editing complexes. In the absence of Dox, tTR-KRAB binds to *TetO* elements imposing a closed heterochromatic state (high H3K9me3/low H3-acetylation) at target sequences upon the recruitment of KAP1 and HP1 amongst other endogenous chromatin remodelling factors. In the presence of Dox, tTR-KRAB does not bind *TetO* permitting the same target sequences to acquire an open euchromatic state (low H3K9me3/high H3-acetylation). Once the different DNA editing processes are finished, Dox is added for determining the frequencies and types of DNA changes via flow cytometry and targeted deep sequencing analyses. **(B)** Overview of specific experimental set-ups. The tTR-KRAB-expressing HER.TLR$^{TetO.KRAB}$ cells have a *TetO*-flanked traffic light reporter (*TLR*) containing the *EGFP* reading frame interrupted by heterologous sequences and a stop codon linked to a T2A peptide "self-cleavage" motif and an out-of-frame *mCherry* reporter. Programmed DNA insertions upstream of the stop codon placing the *mCherry* in-frame are measured through mCherry-directed flow cytometry. The control tTR-KRAB-expressing HER.TLR^{KRAB} cells differ from HER.TLR$^{TetO.KRAB}$ cells in that they have a Dox-insensitive *TLR* reporter due to their lack of *cis*-acting *TetO* elements. The *TetO*-flanked *EGFP* construct in tTR-KRAB-expressing HEK.EGFP$^{TetO.KRAB}$ cells is functional with programmed frameshifts and start codon knockouts yielding a traceable EGFP-negative phenotype. DNA editing byproducts disrupting the reading frame or EGFP spectral characteristics (e.g., indels, unintended nucleotide substitutions inside and outside base editing windows, and pegRNA scaffold-derived insertions) contribute to the EGFP-negative cell fraction. In addition, in HEK.EGFP$^{TetO.KRAB}$ cells, precise T-to-C transition events at a specific codon yield a traceable blue light-emitting phenotype resulting from the conversion of the EGFP fluorophore to that of EBFP.

RESULTS

Cell- and DNA-level assays relying on defined epigenetic control over nucleotide target sequences were implemented for quantifying and characterizing genetic changes resulting from the interactions of DNA cutting-free gene editing tools with different higher-order chromatin states (**Figure 1A**). These systems, based on human embryonic kidney cells and retinoblasts (i.e., HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$, respectively), allow live-cell quantification of different DNA editing outcomes (precise or otherwise) through reporter-directed flow cytometry (**Figure 1A**). The precision of the different DNA editing processes can be further analysed via genotyping assays based on deep next-generation sequencing **(Figure 1A**). In brief, in the absence of doxycycline (Dox), target sequences are embedded in heterochromatin impinged by the KRAB-mediated recruitment of KRAB-associated protein 1 (KAP-1) and heterochromatin protein-1 (HP-1) amongst other factors. This compact chromatin state is characterized by high and low deposition of specific epigenetic marks, e.g., H3K9me3 and pan H3-acetylation, respectively[29]. Conversely, in the presence of Dox, the same target sequences are placed in relaxed euchromatin characterized instead by low and high accumulation of H3K9me3 and H3-acetylation, respectively[29].

We started by transfecting HER.TLR$^{TetO.KRAB}$ and control HER.TLR^{KRAB} cells, cultured in the presence or in the absence of Dox (**Figure 1B** and **Supplementary Figure S3**), with expression plasmids expressing PE2 or PE3 components (**Figure 2A**). HER.TLR$^{TetO.KRAB}$ and HER.TLR^{KRAB} cells differ from each other in that the latter lacks *cis*-acting *TetO* elements and, as a result, target sequences retain an euchromatic character with and without Dox (**Figures 1A**). Parallel cell cultures transfected with constructs expressing control reagents (i.e., Cas9^{H840A} and non-targeting gRNA^{NT}), served as negative controls. After the action of the various complexes, all cell cultures were exposed to Dox to allow for prime editing quantification by mCherry-directed flow cytometry (**Figure 1B** and **Supplementary Figure S3**). The resulting data revealed that, in HER.TLR$^{TetO.KRAB}$ cells, prime editing activities induced by PE2 and PE3 complexes were readily detected at target sequences embedded in euchromatin whereas that was not the case at the same target sequences located in heterochromatin (**Figure 2B**, left panel). Importantly, there were no statistically significant differences in prime editing frequencies in control HER.TLR^{KRAB} cells whose target sequences are not under KRAB-dependent epigenetic control (**Figure 2B**, right panel). Similar experiments performed with another set of PE2 and PE3 reagents (**Figure 2C**) in HEK.EGFP$^{TetO.KRAB}$ cells led to results consistent with those obtained in HER.TLR$^{TetO.KRAB}$ cells (**Figure 2B**, left panel). Specifically, prime editing activities at various heterochromatic target sites were significantly lower than those attained at their euchromatic counterparts with, in fact, one of the PE2 complexes (i.e., PE2:pgRNA.31) failing to trigger prime editing above background levels at heterochromatin (**Figure 2D**). As a consequence, in HEK.EGFP$^{TetO.KRAB}$ cells, the ratios between prime editing levels corresponding to epigenetically open versus closed DNA, herein named the chromatin impact index, varied substantially, i.e., from 2- to 5.3-fold (**Figure 2E**).

To assess the relationship between prime edits and prime editing bystander events directly in living cells, we tested PE2 and PE3 complexes containing pegRNA.16 designed to change the EGFP fluorophore to that of EBFP (**Figure 1B** and **Figure 3A**). In line with the previous data (**Figure 2B**, left panel and **Figure 2D**), prime editing with these additional reagents was highest at euchromatin (**Figure**

3B, top panel; and **Figure 3C**) that, as a result, invariably led to chromatin impact indexes superior to 1 (**Figure 3D**, top panel). Notably, this enhanced prime editing activity at euchromatin was accompanied by a significant increase in gene knock-out byproducts in cells treated with the PE3 reagents (**Figure 3B**, bottom panel; and **Figure 3C**), resulting in chromatin impact indexes higher than 1 for these unintended bystander events (**Figure 3D**, bottom panel). The higher prime editing activity of PE2:pegRNA.16 complexes at euchromatin over heterochromatin was confirmed by NGS analysis (**Figure 3E**). In addition, NGS analysis identified indels at the heterochromatic and euchromatic forms of the PE2:pegRNA.16 target site as well as pegRNA scaffold-derived insertions at the latter form (**Figure 3E**).

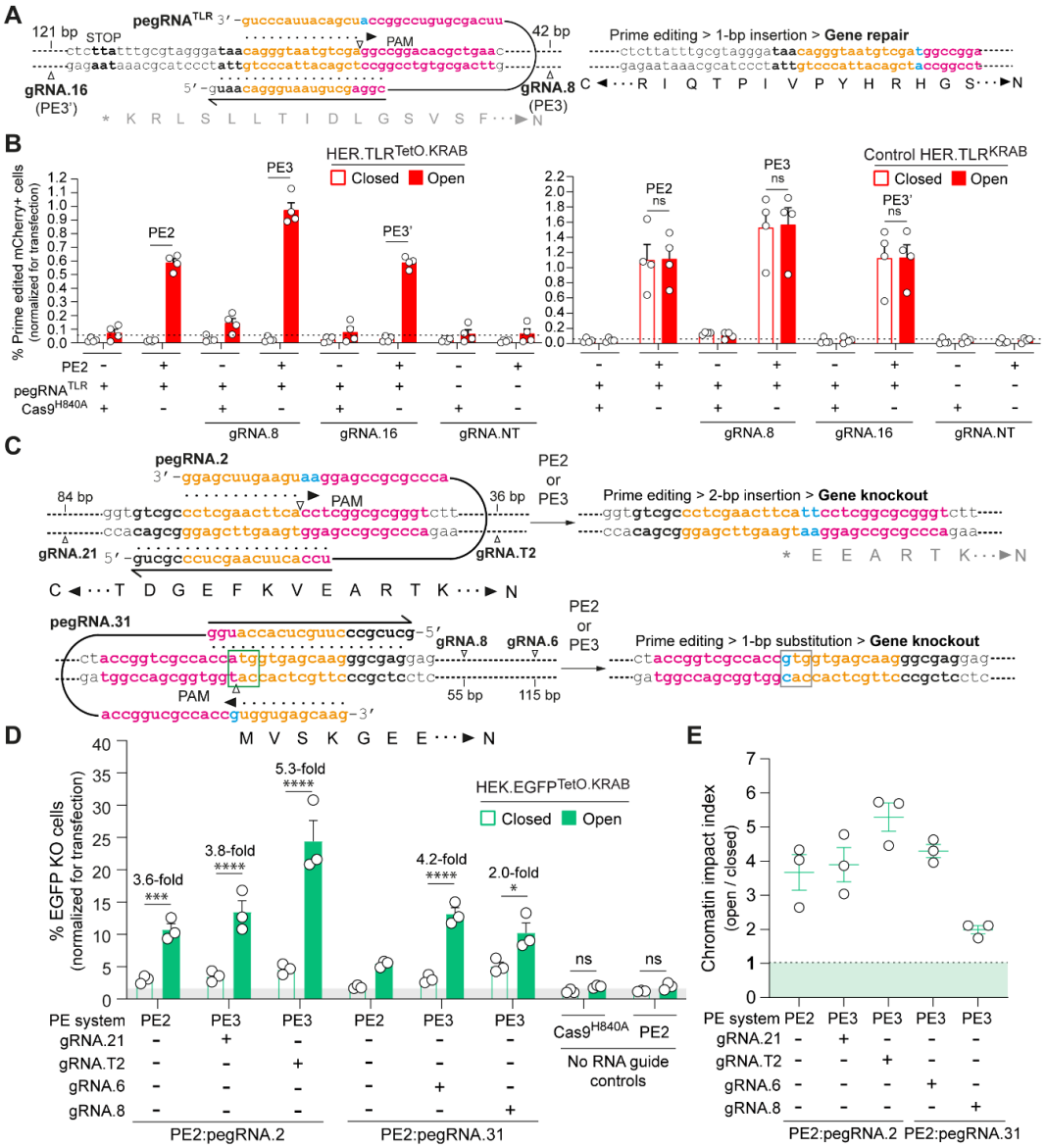


Figure 2. Prime editing at euchromatin versus heterochromatin using gene repair and gene knockout assays. (A) Gene repair set-up. Schematics of target site before and after prime editing with PE2:pegRNA^{TLR} complexes alone (PE2 system) or together with an auxiliary gRNA (PE3 system). Distances (in bp) between nicks defined by pegRNA^{TLR} and each auxiliary gRNA are specified. pegRNA^{TLR} is built to correct the *mCherry* reading frame by inserting a 1-bp (cyan nucleotides). **(B)** Quantification of prime editing in human embryonic retinoblasts. HER.TLR$^{TetO.KRAB}$ and control HER.TLR^{KRAB} cells (left and right panel, respectively), treated and not treated with Dox, received the indicated prime-editing and control reagents. mCherry-directed flow cytometry after sub-culturing and Dox addition establishes prime editing frequencies. Bars and error bars correspond to mean ± s.e.m., respectively (n=4 biological replicates); $P > 0.05$ considered non-significant (ns). (**C**) Gene knockout set-ups. Schematics of pegRNA.2 and pegRNA.31, and their respective target sites, before and after prime editing. Distances (in bp) between nicks defined by pegRNAs and pairing gRNAs for PE3-based DNA editing are shown. pegRNA.2 and pegRNA.32 are designed for target gene knockout through the installation of a 2-bp insertion and a 1-bp substitution that disrupts the reading frame and start codon, respectively (cyan nucleotides). **(D)** Quantification of prime editing in human embryonic kidney cells. HEK.EGFP$^{TetO.KRAB}$ cells, cultured with or without Dox, received the indicated prime-editing and control reagents. Flow cytometry upon sub-culturing and Dox addition established EGFP knockout frequencies. Bars and error bars represent mean ± s.e.m., respectively (n=3 biological

replicates). Significances derived from two-way ANOVA followed by Šídák's test for multiple comparisons; *0.01 < *P* < 0.05; ***0.0001< *P* < 0.001; ****P < 0.0001; P > 0.05 was considered non-significant (ns). **(E)** Relative prime editing activities at open versus closed chromatin. Chromatin impact indexes for the indicated PE2 and PE3 reagents corresponding to the ratios between the mean EGFP knockout frequencies measured in the presence and absence of Dox. Scatter plot displays mean ± s.e.m. (n=3 biological replicates).

Delivery of PE3 RNA reagents together with $Cas9^{H480A}$, instead of PE2, also yielded a significant increase in byproduct accumulation at euchromatin (**Figure 3B**, bottom panel). This data supports the conclusion that most PE3-induced mutagenic events arose from offset nicking at both DNA strands and that these mutagenic events were most prevalent at euchromatin. Interestingly, the 6.3-fold higher prime editing activities at heterochromatin using PE2, pegRNA.16 and gRNA.2, instead of PE2 and pegRNA.16 (**Figure 3B**, top graph), was not accompanied by a significant increase in byproduct build-up (**Figure 3B**, bottom graph). This data suggests that, at certain heterochromatic sequences, judicious selection of auxiliary gRNAs can lead to efficient PE3-mediated editing without a concomitant build-up of DSB-derived indels. Taken together, these data indicate that prime editing is hindered at KRAB-impinged heterochromatin in a target site- and PE system-independent manner.

Next, we sought to probe the capacity of pegRNAs to engage different chromatin states by coupling them to Cas9. To this end, $HER.TLR^{TetO.KRAB}$ and $HER.TLR^{KRAB}$ cells, cultured with or without Dox, were exposed to Cas9 together with $gRNA^{TLR}$ or $pegRNA^{TLR}$ whose spacer and scaffold sequences are the same. At both chromatin states, reading frame repair resulting from Cas9-induced indels was lowest in cells receiving $pegRNA^{TLR}$ (**Figure 4A**, left panel). Moreover, when compared to canonical $Cas9{:}gRNA^{TLR}$ complexes, $Cas9{:}pegRNA^{TLR}$ complexes were the most hindered by heterochromatin (**Figure 4A**, left panel), as highlighted by their 5.4-fold higher chromatin impact index (**Figure 4A**, right panel). In control $HER.TLR^{KRAB}$ cells, significant differences in DNA editing events between cultures treated and untreated with Dox were, once again, not detected regardless of the tools used (**Figure 4B**). Similar experiments performed in $HEK.EGFP^{TetO.KRAB}$ cells with another set of gRNAs and pegRNAs sharing the same spacers, yielded outcomes consistent with those obtained in $HER.TLR^{TetO.KRAB}$ cells (**Figure 4C**). Firstly, at both chromatin states, gene knockouts resulting from Cas9-induced indels were lowest in cells receiving pegRNAs instead of gRNAs (**Figure 4C** and **Figure 4D**). And, secondly, when compared to canonical Cas9:gRNA complexes, Cas9:pegRNA complexes were the most impeded by heterochromatin (**Figure 4C**), as underscored by their 2-fold higher chromatin impact indexes (**Figure 4E**). Taken together, these data shows that pegRNAs can contribute to the underperformance of prime editing complexes at KRAB-regulated heterochromatin.

Tethering KRAB domains to chromosomal sequences through DNA-binding motifs of native and engineered proteins (e.g., zinc-finger-KRAB and tTR-KRAB proteins, respectively) can locally nucleate *bona fide* heterochromatin[29,33-35]. However, the resulting heterochromatin is not maintained if pioneering KRAB-containing proteins are solely recruited. Thus, clearly, additional epigenetic factors are necessary for depositing specific combinations of DNA methylation and histone modifications that, together, underpin stable and heritable heterochromatic states. Importantly, a single fusion protein named CRISPRoff consisting of a catalytically "dead" Cas9 scaffold linked to KRAB and to two DNA methyltransferases (i.e., DNMT3A and DNMT3L), has recently been shown to assemble stable heterochromatin through RNA-programmable binding to endogenous gene control regions[36]. Hence, to further study the role of chromatin controlled by specific epigenetic factors on prime editing, we next applied the CRISPRoff system to epigenetically remodel *CD81* alleles. To this end, after exposing HEK293T cells to CRISPRoff and gRNAs targeting *CD81* regulatory sequences (**Supplementary Figure S4**), cells acquiring a $CD81^-$ phenotype were sorted from their $CD81^+$ counterparts via FACS (**Figure 5A**). The sorted $CD81^-$ and $CD81^+$ cell populations kept their respective phenotypes upon long-term culturing (**Figure 5B**). Importantly, as demonstrated by ChIP-qPCR analyses, $CD81^-$ and $CD81^+$ cells contained *CD81* alleles with epigenetic marks characteristic of heterochromatin and euchromatin, respectively. Specifically, *CD81* sequences in $CD81^-$ cells were enriched in histone 3 lysine 9 trimethylation (H3K9me3) and depleted in histone 3 acetylation (H3Ac) (**Figure 5C**). Conversely, *CD81* sequences in $CD81^+$ cells were depleted in H3K9me3 and enriched in H3Ac (**Figure 5C**). The $CD81^-$ and $CD81^+$ cell populations were transfected with constructs expressing prime editing complexes (n=10), designed to install 1-bp substitutions at epigenetically silenced and active *CD81* alleles, respectively (**Figure 5D**). Western blot analysis with a Cas9-specific antibody confirmed similar prime editor expression levels (**Figure 5E**), and the absence of otherwise interfering CRISPRoff complexes in

the CD81⁺ and CD81⁻ cells (**Figure 5E**, Mock lanes). Deep sequencing analysis revealed that amongst the ten prime editing complexes assembled, seven yielded higher prime editing activities in CD81⁺ than in CD81⁻ cells (up to 2.8-fold); with statistical significance being reached in cells exposed to five of these complexes (**Figure 5F**). In contrast, the higher prime editing activities measured in CD81⁻ than in CD81⁺ cells resulting from the three additional prime editing complexes did not reach statistical significance (**Figure 5F**). Taken these data together, we conclude that prime editing is mostly unfavoured at heritable heterochromatin controlled by the combined recruitment of KRAB and DNA methyltransferases DNMT3A and DNMT3L.

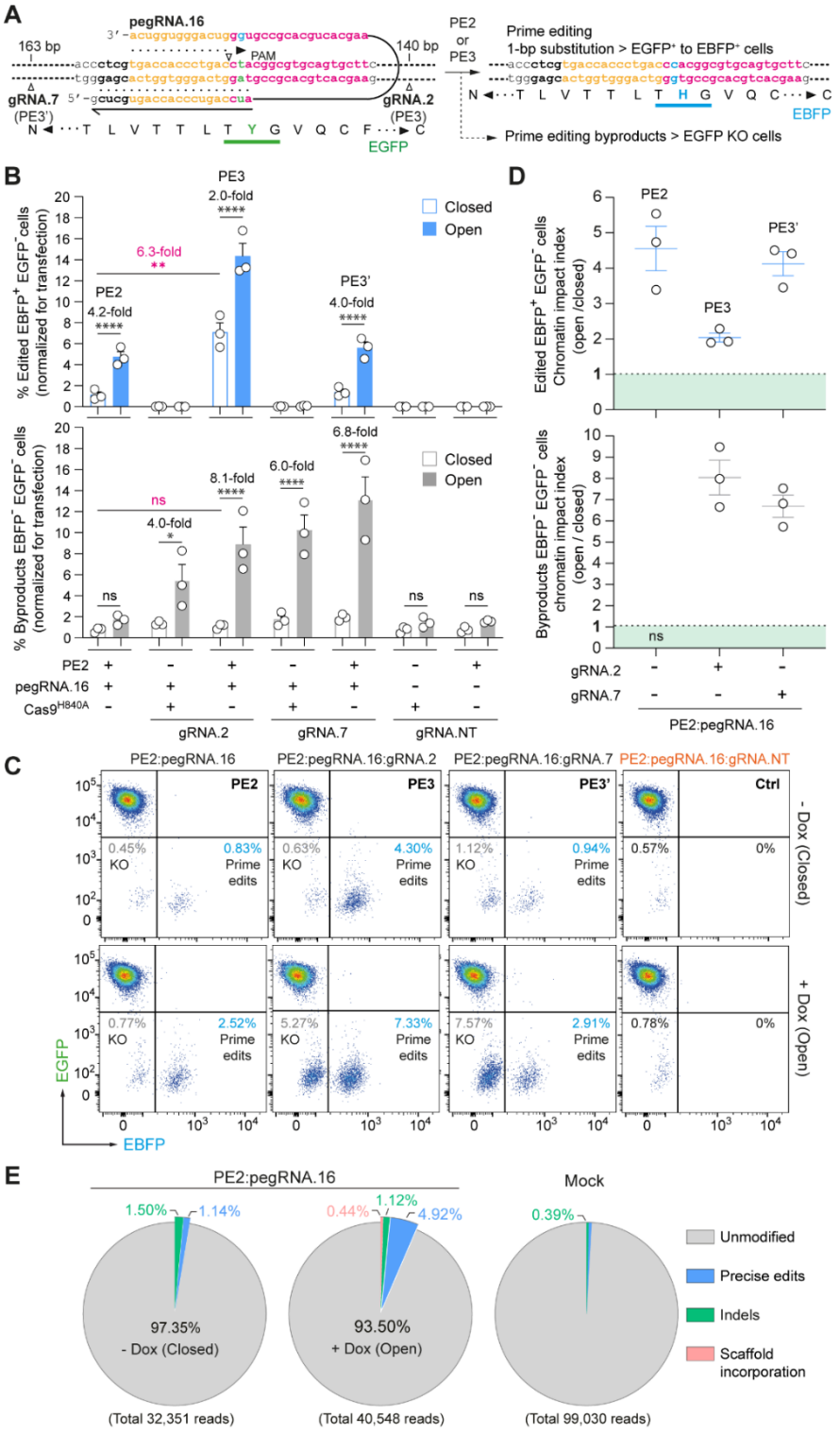


Figure 3. Prime editing outcomes at euchromatin versus heterochromatin combining live-cell gain and loss of function assays. (A) Prime editing set-up. Schematics of pegRNA.16 and cognate target sequence before and after prime editing with PE2:pegRNA.16 complexes alone (PE2 system) or together with an auxiliary gRNA (PE3 system). Distances (in bp) between nicks defined by pegRNA.16 and each auxiliary gRNA are specified. pegRNA.16 is designed to change the fluorophore of EGFP to that of EBFP (underlined residues). **(B)** Flow cytometric quantification of prime editing outcomes. HEK.EGFP$^{TetO.KRAB}$ cells, treated or not treated with Dox, were exposed to the indicated prime-editing or control components. Flow cytometry after sub-culturing and Dox supplementation quantified prime editing (i.e., EBFP-positive cells) and gene knockout by-product events (i.e., EBFP/EGFP doubly negative cells). Bars and error bars represent mean ± s.e.m., respectively (n=3 biological replicates). Significance between - and + doxycycline datasets derived from two-way ANOVA followed by Šídák's test for multiple comparisons; *0.01 < *P* < 0.05; ****P < 0.0001; P > 0.05 considered non-significant (ns). Significance between the indicated - doxycycline datasets was determined with two-tailed Student's t test; **0.001 < *P* < 0.01; P > 0.05 considered non-significant (ns). **(C)** Representative dot plots corresponding to the experimental results plotted in panel B. **(D)** Relative prime editing outcomes at open versus closed chromatin. Chromatin impact indexes corresponding to prime editing and indel-derived gene knockout events at open and closed chromatin (top and bottom panel, respectively) induced by the indicated components. Scatter plots display mean ± s.e.m. (n=3 biological replicates). (**E**) Deep sequencing quantification of prime editing outcomes. Pie chart parsing the frequencies of prime editing events in HEK.EGFP$^{TetO.KRAB}$ cells treated and untreated with Dox and exposed to PE2:pegRNA.16 complexes. Mock-transfected cells served as negative control.

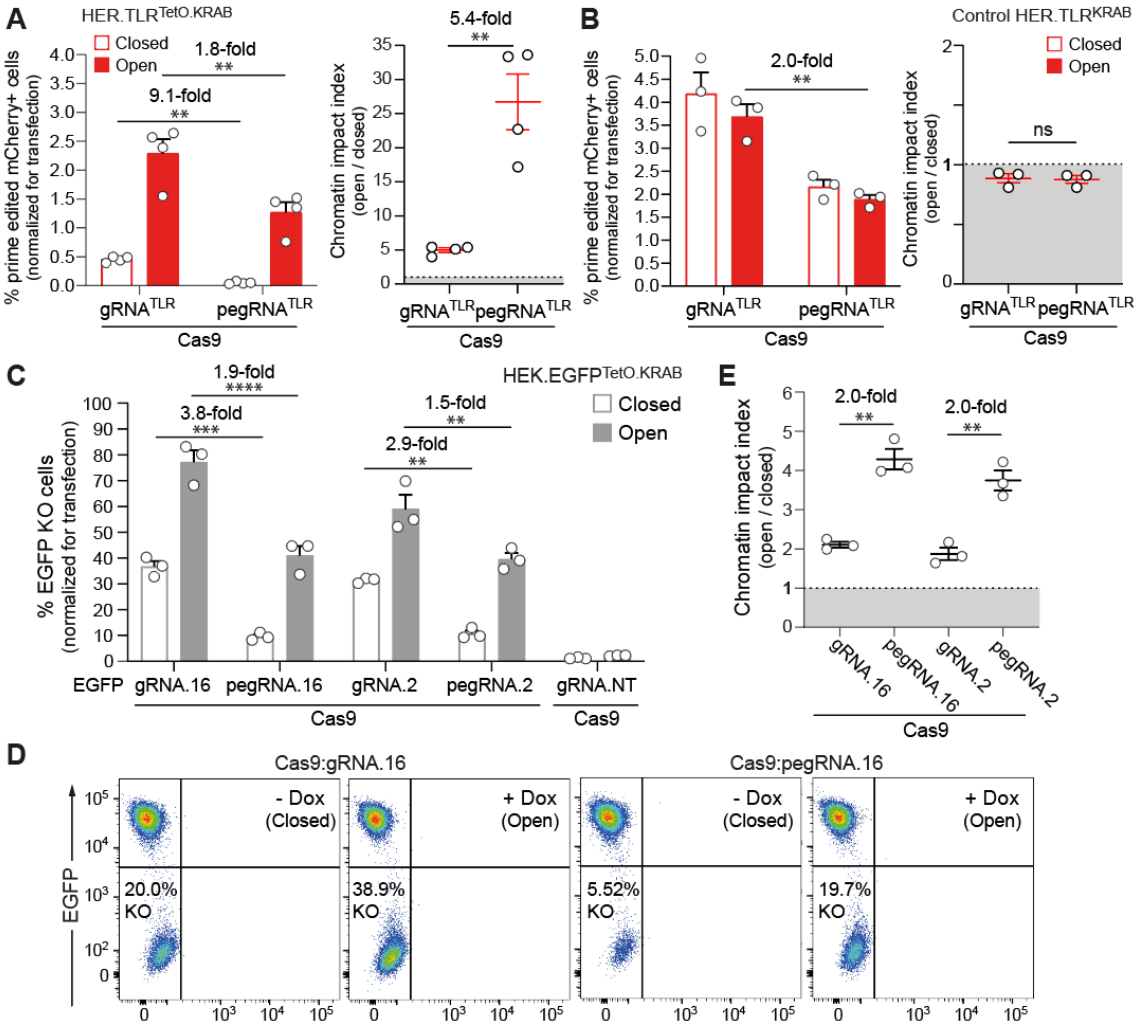


Figure 4. Probing the performances of gRNAs versus pegRNAs at euchromatin and heterochromatin. (**A**) Gain-of-function assessments in HER.TLR$^{TetO.KRAB}$ cells. Cells treated and untreated with Dox received either canonical Cas9:gRNA^{TLR} or Cas9:pegRNA^{TLR} complexes. After sub-culturing and Dox supplementation, flow cytometry quantified Cas9-mediated correction of the *mCherry* reading frame. Gene repair frequencies and corresponding chromatin impact indexes are plotted. (**B**) Gain-of-function assessments in control HER.TLR^{KRAB} cells. The same procedures and reagents applied to HER.TLR$^{TetO.KRAB}$ were also used in *TetO*-negative HER.TLR^{KRAB} control cells. Gene repair frequencies and corresponding chromatin impact indexes are plotted. (**C**) Loss-of-function assessments in HEK.EGFP$^{TetO.KRAB}$ cells. Cells cultured with and without Dox, were exposed to the indicated reagents. Flow cytometry after sub-culturing and Dox addition established Cas9-induced EGFP knockout frequencies. (**D**) Representative dot plots corresponding to experimental results shown in panel C. (**E**) Relative gene knockout levels at open versus closed chromatin. Chromatin impact indexes for the indicated reagents were assessed by calculating the ratios between the mean EGFP knockout levels measured in the presence and absence of Dox. Results are depicted as mean ± s.e.m. of 3 or 4 independent biological replicates. Significances amongst gene repair and gene knockout datasets were calculated via two-way ANOVA followed by Šídák's test for multiple comparisons. Significances between chromatin impact index datasets were

determined with unpaired two-tailed Student's t test. **0.001 < P < 0.01; ***0.0001< *P* <0.001; ****$P < 0.0001$; $P > 0.05$ considered non-significant (ns).

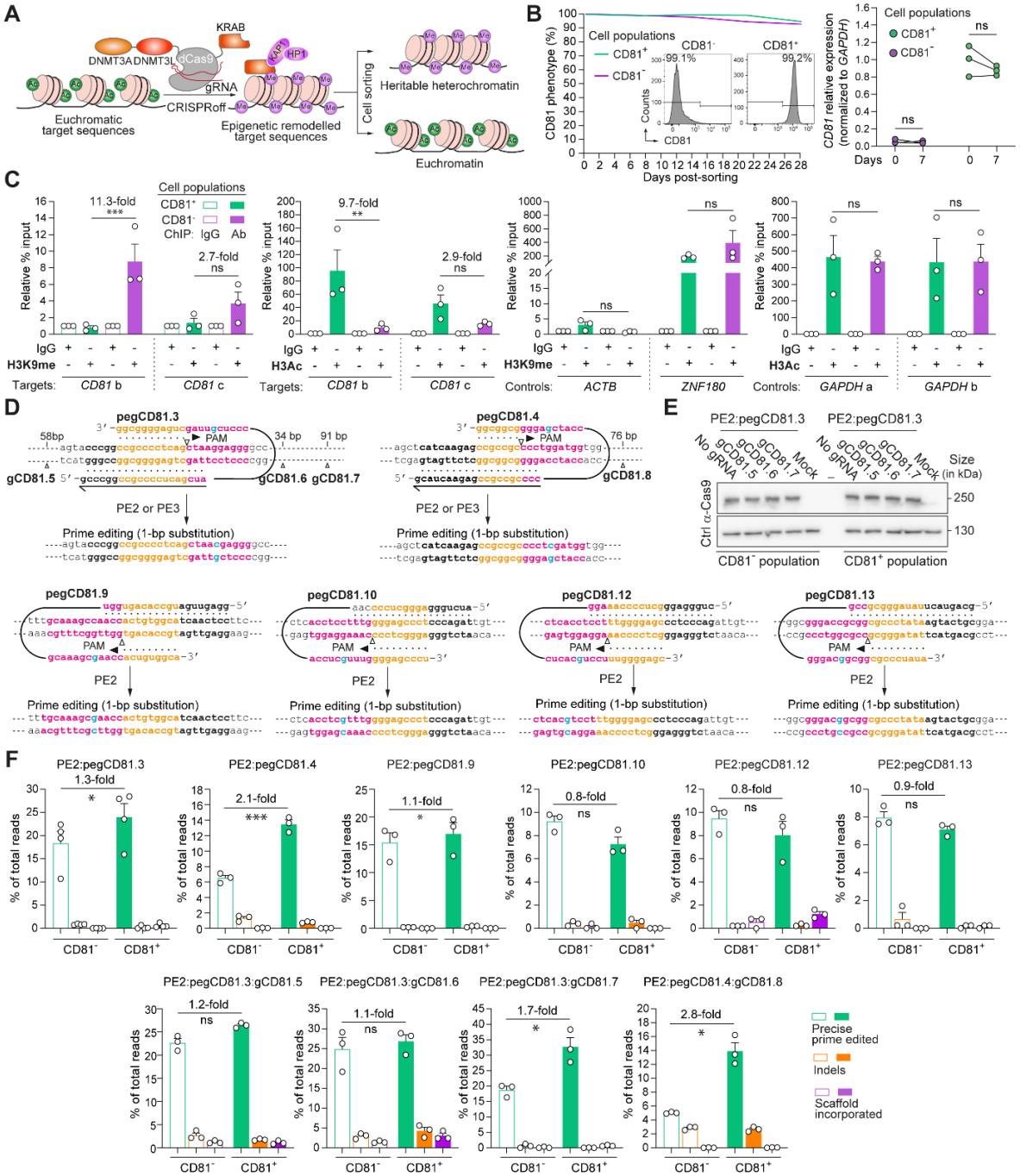


Figure 5. Prime editing at open and heritable closed chromatin. (**A**) Illustration of experimental set-up. HEK293T cells with *CD81* in epigenetically open and closed states were generated through CRISPRoff transfection and CD81-based cell sorting (CD81$^+$ and CD81$^-$ populations, respectively). (**B**) Validation of stable CD81 phenotypes. CD81$^+$ and CD81$^-$ phenotypes are maintained upon sub-culturing as assessed through flow cytometry and RT-qPCR analyses (left and right panel, respectively). Inset, representative flow cytometry histogram. Significances were calculated via two-way ANOVA followed by Šídák's test for multiple comparisons. $P > 0.05$ considered non-significant (ns). (**C**) Validation of stable *CD81* epigenetic states. ChIP-qPCR analyses of two *CD81* regions in CD81$^+$ and CD81$^-$ cells presenting preferential enrichment of the heterochromatin mark H3K9me3 in the former; and the euchromatin mark pan-H3 acetylation (H3Ac) in the latter. ChIP-qPCR analyses controls involved assessing H3K9me3 and H3Ac marker deposition at loci with known open chromatin (i.e., *ACTB* and *GAPDH*) and closed chromatin (i.e., *ZNF180*) conformations. ChIP-qPCR data, controlled for background (IgG) and normalized for input chromatin, are plotted as mean ± s.e.m. of percentage of input values (n=3 technical replicates). Significances were calculated via two-way ANOVA followed by Tukey's test for multiple comparisons. **0.001 < *P* < 0.01; ***0.0001< *P* <0.001; $P > 0.05$ considered non-significant (ns). (**D**) Prime editing set-ups at *CD81*. Schematics of pegRNA.CD81.1 and pegRNA.CD81.2 and their respective target sites prior to and after prime editing. Distances (in bp) between nicks defined by pegRNAs and each pairing gRNA for PE3-based DNA editing are also depicted. *CD81*-targeting pegRNAs are designed for installing G-to-C substitutions at two separate

places within the gene (cyan nucleotides). (**E**) Controlling chromatin-remodelling and prime-editing complex levels. Western blot analysis confirming the absence of CRISPRoff proteins in long-term cultures of CD81$^-$ and CD81$^+$ cells (Mock) and establishing similar amounts of prime editors in these cultures at 48 h post-transfection. (**F**) Quantification of prime editing at alleles with heritable open and closed chromatin. Targeted deep sequencing analysis of CD81$^-$ and CD81$^+$ cells at 7 days post-transfection of constructs expressing the indicated PE2 and PE3 complexes. Mock-transfected cells served as negative controls. Bars and error bars represent mean ± s.e.m., respectively, of 3 independent biological replicates. Significances between indicated datasets were determined with paired two-tailed Student's t test. *0.01 < *P* < 0.05; ***0.0001< *P* <0.001; *P* > 0.05 considered non-significant (ns).

To start investigating the performance of base editors at euchromatin versus heterochromatin, HEK.EGFP$^{TetO.KRAB}$ cells cultured with or without Dox, were transfected with constructs expressing CBE or ABE complexes designed for *EGFP* knockout through site-directed mutagenesis (**Supplementary Figure S5**). Strikingly, flow cytometry analysis readily disclosed that, when compared to prime editing complexes, base editing complexes are typically the least affected at heterochromatin (**Figure 6A**). Indeed, amongst the eight base editors initially assembled, six presented either similar efficiencies at euchromatin and heterochromatin or even higher efficiencies at the latter compact DNA state (**Figure 6A**) yielding, as a result, chromatin impact indexes around or under 1, respectively (**Figure 6B**). The fluorophore exchange capacity of ABE:gRNA.32 was further explored to confirm its similar performance at euchromatin and KRAB-regulated heterochromatin via flow cytometric quantification of cells acquiring EBFP-specific fluorescence (**Figure 6C** and **Figure 6D**).

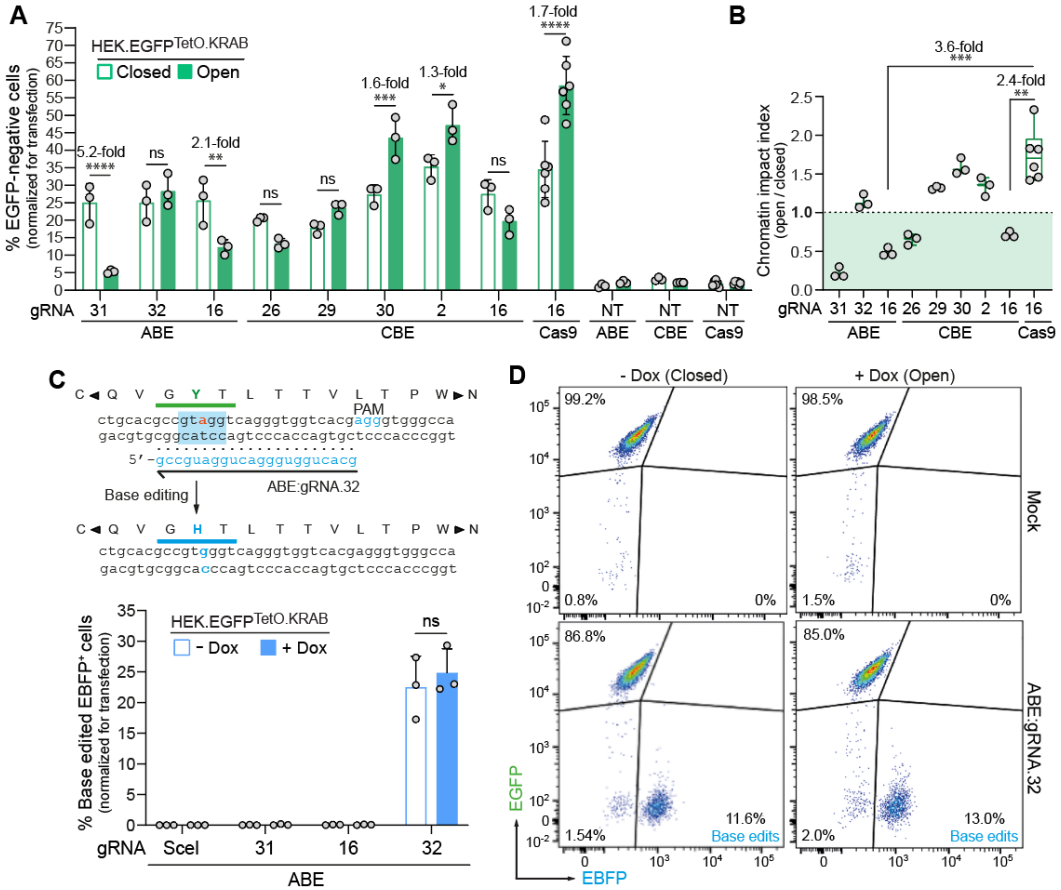


Figure 6. Investigating base editing at euchromatin versus heterochromatin with combined loss and gain of function live-cell assays. (**A**) Gene knockout set-ups. HEK.EGFP$^{TetO.KRAB}$ cells, cultured with or without Dox, were treated with the indicated base-editing and control reagents. Flow cytometry upon sub-culturing and Dox addition established EGFP knockout frequencies. Bars and error bars denote mean ± s.d. of independent biological replicates, respectively. Significances were calculated via two-way ANOVA followed by Šídák's test for multiple comparisons; *0.01 < *P* < 0.05; **0.001 < *P* < 0.01; ***0.0001< *P* < 0.001; ****P < 0.0001; P > 0.05 considered non-significant (ns). **(B)** Relative base editing activities at open versus closed chromatin. Chromatin impact indexes for the indicated base-editing and control reagents correspond to the ratios between the mean EGFP knockout levels measured in the presence and absence of Dox. Scatter plot displays mean ± s.d. Significances were calculated via Student's *t* test; **0.001 < *P* < 0.01; ***0.0001< *P* < 0.001 **(C)** Gene conversion set-ups. Schematics of gRNA.32 and cognate target sequence before and after prime editing with ABE:gRNA.32 complexes. Target and product nucleotides withing the editing window are marked in red and cyan, respectively. gRNA.32 (spacer shown) is designed to change the fluorophore of EGFP to that of EBFP (underlined residues). HEK.EGFP$^{TetO.KRAB}$ cells, treated or not treated with Dox, were exposed to ABE:gRNA.32. EBFP-directed flow cytometry after sub-culturing and Dox addition established base editing frequencies. Significance was assessed by using

two-tailed Student's t test; *P* > 0.05 was considered non-significant (ns). **(D)** Representative dot plots corresponding to experimental data plotted in panel C.

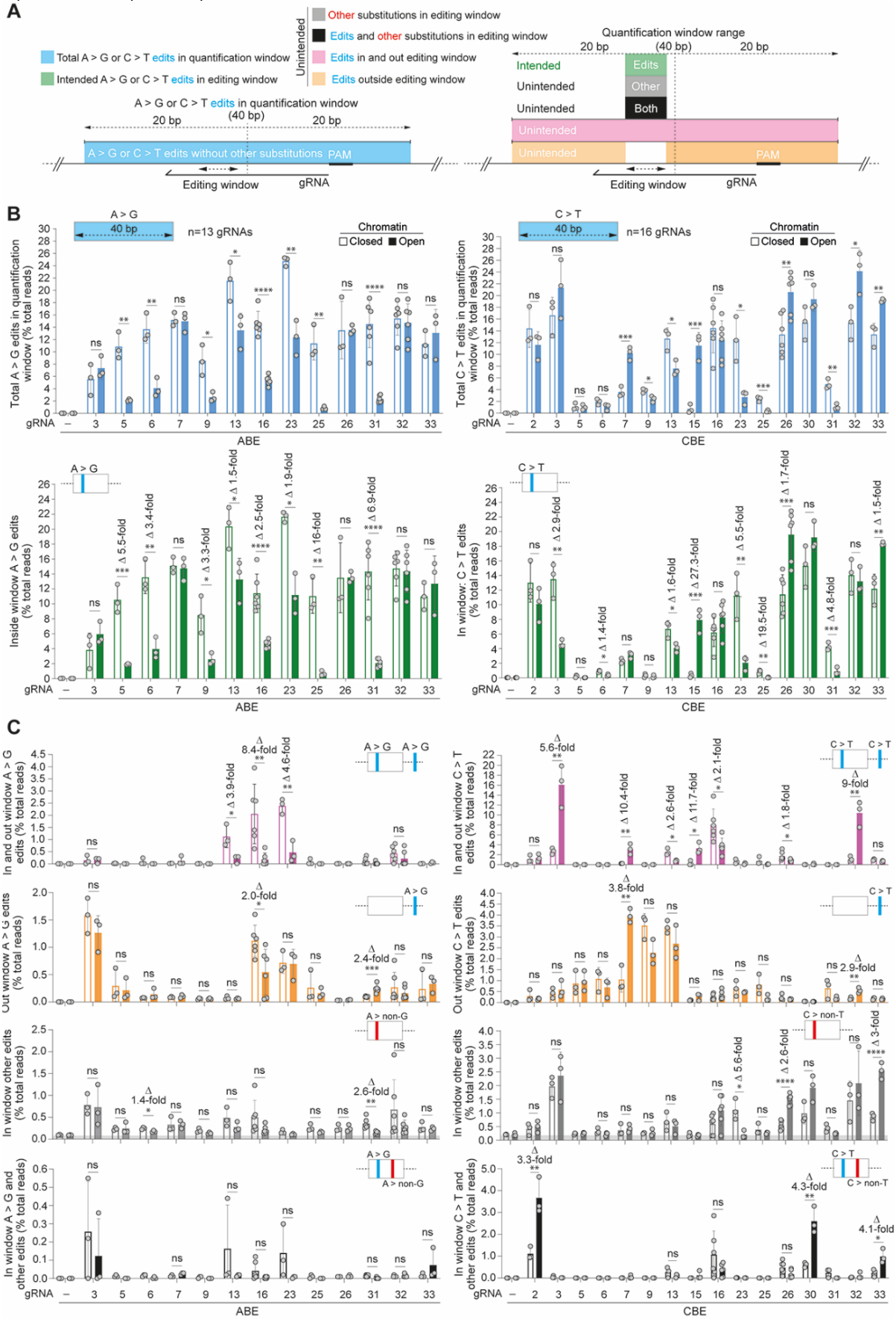


Figure 7. Base editing at open and KRAB-controlled closed chromatin. (**A**) Diagram of the distribution of intended and unintended base-editing products. Base edits (i.e., A > G or C > T) within quantification and editing windows are labelled in cyan and green, respectively. A > G or C > T base edits present simultaneously inside and outside editing windows are marked in magenta and A > G or C > T base edits present exclusively outside editing windows are labelled in orange. Unintended substitutions (i.e., A > non-G and C > non-T) within editing windows (other) are marked in grey. Composite edits consisting of intended and unintended substitutions inside editing windows are labelled in black. (**B**) Quantification of base edits at open and KRAB-controlled closed chromatin. HEK.EGFP$^{TetO.KRAB}$ cells incubated with or without Dox were exposed to the indicated ABE:gRNA and CBE:gRNA complexes designed for installing A > G and C > T substitutions, respectively. Total base edits within the quantification window and intended base edits within each editing window are plotted (top and bottom graphs, respectively). (**C**) Quantification of base-editing byproducts at open and KRAB-controlled closed chromatin. Base-editing byproducts corresponding to base edits inside and outside editing windows or only outside these windows are plotted as magenta and orange bar graphs, respectively. Base-editing byproducts corresponding to unintended substitutions and composite edits consisting of intended and unintended substitutions are plotted as grey and black bar graphs. Base editing events were measured through deep sequencing analyses (50,000 paired-end reads). Significances were determined via two-tailed Student's *t* tests with bars and error bars corresponding to mean ± s.d., respectively (n=3 biological replicates). $P > 0.05$ considered non-significant (ns).

Bystander effects created by ABEs and CBEs include conversion of base pairs outside their predicted editing windows and conversion of target-to-unintended base pairs inside and/or outside those editing windows (i.e., A·T and C·G changing to base pairs other than G·C and T·A, respectively) that, as a consequence, reduce their precision and product purity, respectively (**Figure 7A**). The frequencies and proportions between intended and unintended bystander events and between the different types of the latter products ultimately determines the performance of specific base editing reagents. Notably, whether the epigenetic context of nucleotide sequences affects these key base-editing parameters remains an open question. To address this question, an expanded panel of ABE and CBE complexes, was applied to isogenic target sequences placed in euchromatin versus KRAB-regulated heterochromatin using the HEK.EGFP$^{TetO.KRAB}$ cell system (n=29) (**Figure 7B** and **Supplementary Figure S6**); or embedded in euchromatin versus heritable heterochromatin controlled by the combined actions of KRAB, DNMT3A and DNMT3L using the CD81^{+}/CD81^{-} HEK293T cell system (n=9) (**Supplementary Figures S7** and **S8**). Deep sequencing analyses of A > G and C > T edits within a 40-bp quantification window and within the canonical ABE and CBE base editing windows (**Figure 7A**, cyan and green bins, respectively) confirmed that, diversely from prime editing, base editing activities at heterochromatin are frequently as high as or higher than those registered at euchromatin (**Figure 7B** and **Supplementary Figure S8A**). Indeed, the installation of the intended ABE- and CBE-derived edits was mostly either unhindered (i.e., 5/13 and 6/16, respectively) or in fact fostered (i.e., 8/13 and 5/16, respectively) at heterochromatin, resulting in a majority of base editing complexes presenting a chromatin impact index at or below 1, respectively (**Figure 8**). Remarkably, amongst the ABE and CBE complexes tested, only ABE:gRNA.81.11 (1/17) and 6 CBE:gRNA complexes (6/21), respectively, led to higher base editing at euchromatin than heterochromatin (**Figure 7B** and **Supplementary Figure S8A**).

Further deep sequencing analyses directed at investigating cause-effect relationships between alternate chromatin states and the modulation of base-editing byproducts parsed in 4 different categories (**Figure 7A**, magenta, orange, grey and black bins), showed that the chromatin environment can indeed significantly influence the precision and purity attained by base editors in a gRNA-dependent manner (**Figure 7C** and **Supplementary Figure S8B**) that results in varying proportions between base edits and different types of unintended byproducts (**Supplementary Figure S9**). Interestingly, in these cases, it was observed that diversely from ABE complexes, whose base edits inside and outside editing windows were both mostly higher in heterochromatin, CBE complexes often did not lead to a direct correlation between the frequencies of these substitutions at the two alternate chromatin states. In addition, CBE complexes presented a higher tendency for "spilling over" base editing outside their canonical windows at euchromatin when compared to their ABE counterparts (**Figure 7C**, magenta and orange bars; and **Figure 9**). Finally, consistently with earlier experiments[23], amongst the ABE and CBE complexes tested, the latter were more prone to yielding higher target-to-unintended substitutions than the former (**Figure 7C** and **Supplementary Figure S8B**, grey and black bars). Substitutions reducing CBE product purity have been linked in *Saccharomyces cerevisiae* to the action of specialized DNA polymerases that underpin mutagenic translesion synthesis (TLS)[37]. Our data further discloses that base editing product purity, controlled by TLS or other processes, can vary at alternate chromatin states in a gRNA-dependent manner.

In conclusion, in this study, we demonstrate that the chromatin environment has a significant bearing not only on the activity but also on the precision and product purity attained by DSB-free genome

engineering technologies based on prime editors and base editors. Notably, in striking contrast with programmable nucleases and prime editors, heterochromatin states can in fact favour the activities of base editing complexes.

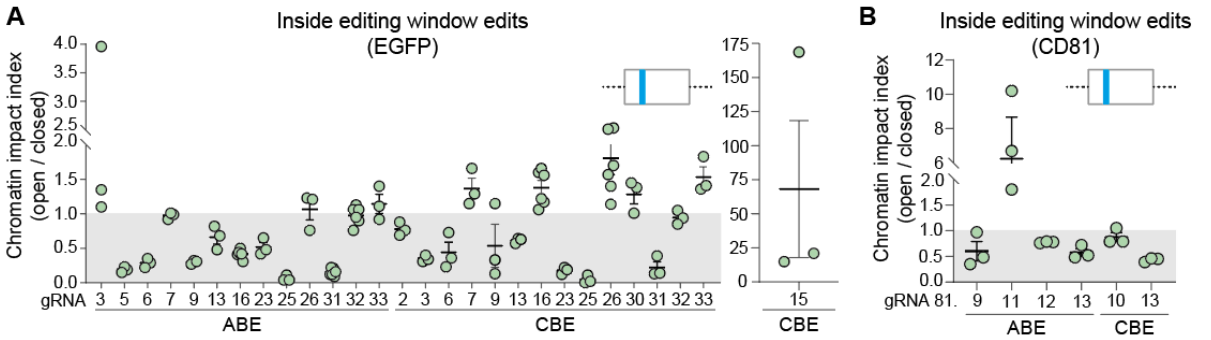


Figure 8. Relative base editing activities at open versus closed chromatin. Chromatin impact indexes for the indicated reagents corresponding to the ratios between base-editing frequencies as determined by deep sequencing at (**A**) *EGFP* and (**B**) *CD81* target sites in open and closed chromatin. Base editing frequencies were determined through amplicon deep sequencing (50,000 paired-end reads). Scatter plot displays mean ± s.d. values.

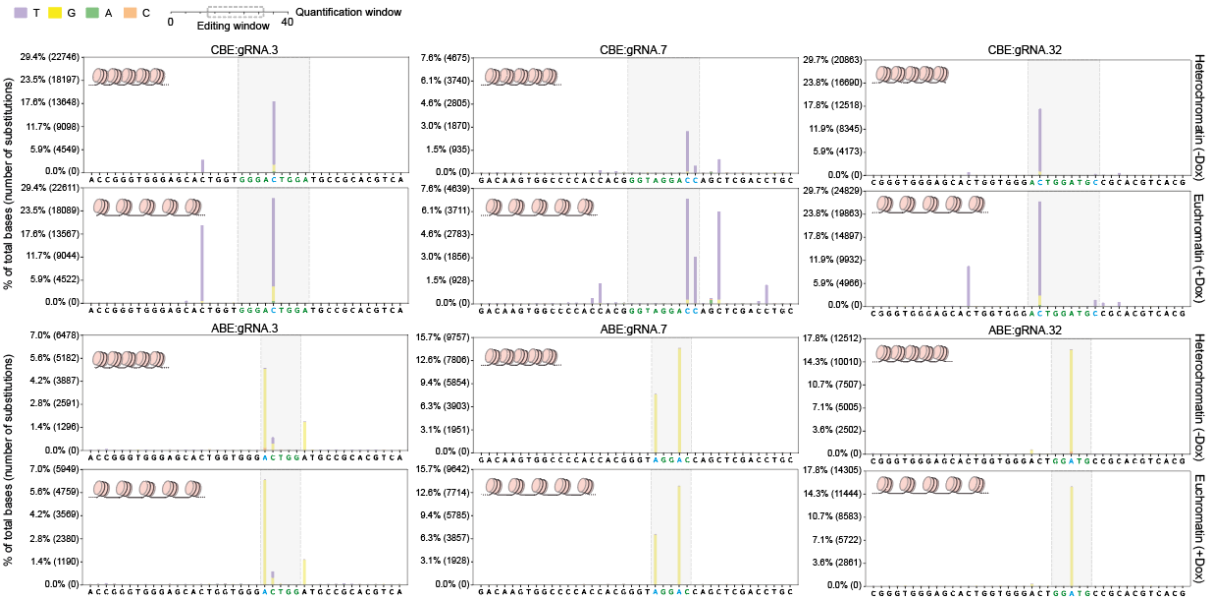


Figure 9. Characterization of base editing events. Representative histograms depicting the type, range and frequency of base editing outcomes generated by the indicated CBE and ABE complexes at open versus closed chromatin in HEK.EGFP$^{TetO.KRAB}$ cells. Green and cyan letters mark the editing window and target nucleotide sequences, respectively, for each of the base editing complexes.

DISCUSSION

The activity of genome-editing reagents is generally dependent on a combination of genetic and epigenetic variables, i.e., nucleotide sequences per se and their epigenetically-regulated chromatin environment, respectively. In this work, to strictly dissect these variables and, as a result, directly investigate the influence of higher-order chromatin conformations not only on the activity but also on the precision of prime editors and base editors, we implemented complementary cellular systems in which isogenic euchromatic and heterochromatic target sites are installed by the recruitment of well-defined epigenetic remodelling factors.

Previous studies from our laboratory and those of others have established that programmable nucleases, including those derived from CRISPR systems, are primarily active at euchromatin than heterochromatin with differential PAM and protospacer accessibility at these distinct higher-order chromatin states constituting a likely determinant factor[29-32]. Similarly to programmable nucleases, recent studies indicate that prime editors are also frequently more active at euchromatic sequences[38,39]. The findings reported here are consistent with these recent studies and further point to the extended 3' ends of pegRNAs (i.e.,

PBS and RT template sequences) as possible culprits for the underperformance of prime editors at heterochromatic sites and for their typically higher chromatin impact indexes than Cas9 nucleases. Of notice, the protection from exonucleolytic degradation of pegRNAs with structured RNA pseudoknots at their 3' ends suggests that extended pegRNA sequences protrude from prime editing complexes[40]. One can postulate that such protruding RNA sequences are more likely to become "trapped" at heterochromatic sites due to local interactions with positively charged deacetylated histone tails and, conversely, become more freely available for DNA flap hybridization and RT-driven cDNA synthesis in euchromatin owing to higher levels of negatively charged acetylated histone tails (**Supplementary Figure S10**). In addition, one can also postulate that DNA flap intermediates emerging during prime editing (**Supplementary Figure S2**), equally associate with deacetylated positively charged histones characteristic of heterochromatic states and, in doing so, contribute to dampened DNA editing. Finally, experiments using PE3 RNA reagents (i.e., pegRNA/gRNA pairs) and the Cas9^{H840A} nickase instead of the whole prime editor protein, confirmed that most mutations caused by PE3 components arise from offset nicking at both DNA chains and further disclosed that these byproducts can build-up at euchromatin.

Experiments correlating CBE BE4[21] and ABE7.10[23] efficiencies with the DNase I hypersensitive site profiling of cognate gRNA target sequences in HEK293T cells suggest that CBE activities are, on average, higher at open than at closed chromatin (1.9-fold); whilst ABE activities are barely affected by chromatin accessibility (1.1-fold). Intriguingly, when compared with their parental proteins, ABE8e-V106W[41] and ABEmax[42] base editors fused to chromatin remodelling and transcription activating domains (i.e., HMGN1 and SOX2 fusions, respectively) yielded lower, similar or higher DNA editing frequencies depending on the loci and nucleotide positions[43,44]. Variable DNA editing by CBE and ABE fusion constructs were also shown to be dependent on the type and fusion location of the effector domains selected on the basis of their involvement in chromatin relaxation[43,44]. Of notice, correlations between histone deacetylase inhibitor (HDACi) treatments and enhanced base editing at certain loci have been interpreted as the result of HDACi-dependent base editor and gRNA expression upregulation and/or increased target sequence accessibility[45,46]. However, in addition to increasing collateral off-target DNA editing[45,46], HDACi treatments have pleiotropic effects that may alter the DNA editing processes themselves via their known modulation of cellular DNA repair pathways, including BER and MMR[47]. This consideration is strengthened by the observation that HDACi treatments while increasing the activity of prime editing complexes designed for installing insertions or deletions, seem to decrease the activity of similar complexes assembled instead for the incorporation of point mutations[46]. Finally, base editors and Cas9 nucleases coupled to the same gRNA can present highly disparate activities[48,49] suggesting that additional mechanisms other than PAM and protospacer accessibility per se, contribute to modulating the efficiency of individual base editing complexes.

Independent cell- and DNA-level assays disclosed that, in striking contrast to prime editors, the activity of base editors are often either unhindered or even fostered when nucleotide sequences transition from euchromatic to heterochromatic states. In particular, heterochromatin installed by the recruitment of the KAP-1 and HP-1 scaffolding KRAB domain alone or together with the DNA methyltransferases DNMT3A and DNMT3L. This finding overtly contrasts with the aforementioned, mostly inhibitory, function of heterochromatin on the activities of CRISPR nucleases and prime editors[29-32, 38,39 and this study].

Base excision repair (BER) is a multi-set process initiated by DNA glycosylases that, through the removal of specific aberrant bases, creates abasic sites that serve as substrates for the apurinic/apyrimidinic endonuclease 1 (APE1). The resulting SSB or gap is ultimately filled-in and sealed by DNA polymerase β and DNA ligase III/XRCC1, respectively. Dissecting the molecular processes underpinning the herein reported surprising finding that base editing can be fostered at heterochromatin will require further research. In this context, it is enticing to postulate a role for heterochromatin in obstructing BER pathways that eliminate CBE and ABE deaminated nucleotides in the form of deoxyuridine and deoxyinosine/hypoxanthine, respectively. Indeed, optimized architectures of CBEs incorporate fusions to an uracil-DNA glycosylase (UDG) inhibitor (UGI) to counteract BER activity and hence improve DNA editing efficiency (**Supplementary Figure S10**). Interestingly, there are also indications that ABE and UGI fusion products can equally foster DNA editing at certain loci[48]. These fusion constructs build on the earlier observation that a class of UDG family members present in fact hypoxanthine-DNA glycosylase activity[50]. Equally consistent with a role for heterochromatin in favouring base editing

through BER inhibition are *in vitro* experiments showing that BER enzymes (e.g., DNA glycosylases, APE1 endonuclease and DNA polymerase β) are substantially more obstructed at chemically modified nucleotides located inward nucleosome surfaces than at the same nucleotides located on more outward positions or naked DNA[51-54]. Significantly, transient DNA unwrapping or directional nucleosome rotations markedly stimulate BER enzymatic activities[51-54]. In line with these results, *in vitro* reconstitution experiments revealed that ATP-dependent chromatin remodelling factors greatly facilitate BER reactions on nucleosome-wrapped DNA[55]. In addition, experiments in cells point to an inverse correlation between the extent of chromatin compaction and BER engagement. In particular, BER complexes assemble preferentially at euchromatic over heterochromatic regions in HeLa cells exposed to base-damaging oxidative stress[56]. It is also noteworthy to mention that, when compared to epigenetically silent loci, open loci greatly facilitate BER initiated by the alkyladenine DNA glycosylase[57], a key enzyme for the removal of bases with aberrant methyl or deoxyinosine/hypoxanthine groups. Collectively, such *in vitro* and *in cellula* datasets suggest that compact chromatin hinders the recruitment and activity of BER enzymes. Hence, in cells, the combined net effects of epigenetically-controlled target DNA accessibility and BER hindrance levels might ultimately determine whether base editing activity is lower, similar or higher at specific sites in open versus closed chromatin (**Supplementary Figure S10**).

Notably, our experiments have further revealed that in addition to base editing frequencies, the buildup of different types of bystander products and their proportions can equally depend on the epigenetic context of target sequences. Bystander effects assessed comprised (i) base editing outside the predefined editing windows of ABE and CBE complexes; and (ii) target-to-unintended base pair conversions. The former and latter byproduct categories define the precision and purity attained by individual base editing complexes, respectively. Besides confirming that ABEs offer substantially higher product purity than CBEs[23], our data suggest that mutagenic translesion synthesis (TLS), a putative source of target-to-unintended CBE products[37], can be contingent upon the epigenetic context of modified bases. Our results further revealed that when compared to ABE complexes, CBE complexes exhibit a higher tendency for base editing "spillover" outside their activity windows upon heterochromatic-to-euchromatic target sequence transitions. As corollary, the activity windows of base editors, in particular CBEs, are not necessarily fixed in that, in addition to specific protein architectures (e.g. effector domains and linker lengths used), they may also depend on extrinsic factors, namely, alternate higher-order chromatin conformations.

In addition to serving as a powerful set of complementary gene-editing tools for basic research, base editors and prime editors constitute high-potential reagents for genetic therapies with base editors having already entered clinical testing[58]. Thus, the performance and safety profiles of these technologies as such or of their individual components in particular (i.e., protein and sequence-specific RNA moieties) necessitates in-depth scrutiny for guiding their selection, further development and application in specific contexts. On the basis of our results, we submit that it will be critical to assess the role of higher-order chromatin environments on the performances of prime editors and base editors as these environments, at both on- and off-target sequences, might vary in different cell types or during the dynamic epigenetic regulation underlying organismal development and cellular differentiation. Furthermore, algorithms trained to predict the activities of DSB-free gene editing reagents, besides target nucleotide sequences per se, will equally profit from processing information on the epigenetic context of said sequences. Finally, our data can further guide the development of combinatorial approaches whereby targeted epigenetic modulators and DSB-free DNA editing tools work in concert for attaining more efficient and/or more precise genomic modifications.

Materials and Methods

Cells

The generation and characterization of the reporter cells HEK.EGFP$^{TetO.KRAB}$ were detailed elsewhere[29]. These cells were kept in high-glucose Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific; Cat.No.: 41966-029) containing 10% fetal bovine serum (FBS; Biowest; Cat.No.: S1860-500). The generation of the human embryonic retinoblasts HER.TLR$^{TetO.KRAB}$, and of their control *TetO*-negative counterparts HER.TLR^{KRAB}, has also been described before[29]. These cells contain the Traffic Light Reporter (TLR) system[59], and were cultured in DMEM supplemented with 10% FBS and 10 mM $MgCl_2$. The human embryonic kidney 293T (HEK293T) cells (obtained from the American Type Culture

Collection), and their CD81-negative and CD81-positive derivatives obtained by FACS after CRISPRoff-mediated *CD81* silencing were cultured in DMEM containing 10% FBS. All cells used in this study were tested for mycoplasma and were kept in a humidified-air 5% CO_2 atmosphere at 37°C.

Recombinant DNA
Standard molecular cloning methods were applied in this study. The gRNA-expressing constructs were assembled by inserting annealed oligonucleotide pairs listed in **Supplementary Table S1** into BveI-digested AY56_pU6.opt-sgRNA.BveI-stuffer[60] or AZ64_pU6.opt-sgRNA.BveI-stuffer. The generation of pegRNA-expressing constructs was initiated by annealing the corresponding oligonucleotide pairs whose sequences are listed in **Supplementary Table S2**. Afterwards, the annealed oligonucleotide pairs were ligated to BsaI-digested AJ71_pU6.PEgRNA-GG-acceptor (Addgene plasmid #132777). The construct gRNA_GFP-T2[61], herein named AT44_gRNA$^{GFP.T2}$, was obtained from Addgene plasmid #41820. Plasmid AM51_pU6.gRNA-I-SceI[29] encoding an irrelevant, non-targeting, gRNA served as a negative control.

Cell transfections
Cell transfections were carried out using as transfection agent 25-kDa linear polyethyleneimine (PEI; Polysciences; Cat.No.: 23966-1) solution (pH 7.4). Prior to transfection, HEK.EGFP$^{TetO.KRAB}$ cells were cultured for 7 days in the presence or absence of 200 ng ml^{-1} doxycycline (Dox; Clontech; Cat. No.: 8634-1). HER.TLR$^{TetO.KRAB}$ and HER.TLR^{KRAB} cells were instead kept for 10 days in medium with or without 500 ng ml^{-1} Dox prior to transfection. Next, the cells were seeded in the indicated culture vessels (**Supplementary Tables S3**-**S16**) and 16 h to 24 h later, transfections were initiated by mixing the relevant plasmids in a 150 mM NaCl solution to which the appropriate amount of a 1 mg ml^{-1} PEI solution was immediately added. After vigorous shaking with a vortex for about 10 sec, the transfection mixtures were incubated for 15 min at room temperature (RT) with the formed DNA-PEI complexes being then directly added into the culture media of the target cells. At 6 h post-transfection, the transfection media were replaced with regular culture media. The cell numbers, the compositions of each DNA mixture used in the different transfection reactions (in ng), the volumes of 150 mM NaCl and PEI solutions (in μl) are specified in **Supplementary Tables S3**-**S16**.

Flow cytometry analyses
The live-cell quantification of gene-editing events resulting from base editing and prime editing was done by using reporter-directed flow cytometry at the indicated timepoints. The initial transfection efficiencies were measured on a per sample basis at 3 days post-transfection for endpoint normalization of gene-editing frequencies. In brief, transfected cells were washed with phosphate-buffered saline (PBS; pH 7.4; Fresenius Kabi; Cat. No.: 16QI2226) and were then treated with a trypsin-EDTA solution (Thermo Fisher Scientific; Cat. No.: 15400-054) to generate single-cell suspensions. After cell collection by a 5-min centrifugation at 300 ×g, the dissociated cells were resuspended in PBS containing 0.5% bovine serum albumin (BSA; Sigma; Cat. No.: A9647-100G) and 2 mM EDTA (pH 8.0). The detection of fluorescence signals was carried out with the aid of a BD LSR II flow cytometer (BD Biosciences) using the appropriate filters. Background fluorescence thresholds were established by applying parental non-transfected cells as negative controls. At least 10,000 viable single cells were acquired per sample. FlowJo 10.6.0 software (Tree Star) was used for data analyses.

***CD81* epigenetic remodeling**
The CRISPRoff system[36] was exploited to investigate the impact of alternate chromatin states on base-editing and prime-editing activities at the endogenous *CD81* locus. The implementation of the CRISPRoff system was initiated by plating HEK293T cells in wells of 24-well plates at a density of 2.0 x 10^5 cells per well. Approximately 17 h later, the cells were transfected by using PEI with the plasmid mixtures indicated in **Supplementary Table S10** designed for *CD81*-targeted epigenetic silencing. After a 7-day subculturing period, the efficiency of targeted gene silencing induced by CRISPRoff was determined by CD81-directed flow cytometry. In brief, cells were washed with PBS (pH 7.4), and subsequently incubated in trypsin-EDTA to generate single-cell suspensions. After trypsin neutralization, 1×10^5 cells were centrifuged at 300 ×g for 5 min after which the pelleted cells were resuspended in 100 ml of ice-cold PBS supplemented with 2% BSA and 2 ml of phycoerythrin (PE)-conjugated anti-CD81 antibody (BD Pharmingen™; Cat. No.: 555676). The cells were stained on ice for at least 30 min in the dark before being washed thrice with ice-cold PBS containing 1% BSA. The frequencies of CD81-

negative cells were determined via a BD LSR II flow cytometer. Next, the cell populations transfected with the CRISPRoff and gRNA plasmid combination yielding the most robust *CD81* silencing levels were expanded. Next, an Arial III flow cytometer was used to sort CD81-negative and CD81-positive populations. CD81 staining and flow cytometry were also carried out for assessing the phenotypic stability of the CD81-negative cell population at 1, 2, 3, and 4 weeks post-sorting. The transfections of constructs encoding prime-editing and base-editing reagents designed for installing 1-bp substitutions at different *CD81* positions, were performed side-by-side in CD81-negative and CD81-positive cells (**Supplementary Tables S11** and **S16**). The cells were then harvested at 2 days post-transfection for western blot analysis of gene-editing proteins and, at 7 days post-transfection, for amplicon deep sequencing analyses of base editing and prime editing in CD81-negative and CD81-positive HEK293T cells. In parallel, RT-qPCR analysis was used for tracing *CD81* mRNA levels in CD81-negative and CD81-positive HEK293T cell populations.

Western blot analyses

The tracing of prime editor proteins in CD81-negative and CD81-positive HEK293T cells was assessed by western blotting. Briefly, at 48 h post-transfection, cells were directly collected in Laemmli buffer consisting of 8.0% glycerol, 3% sodium dodecyl sulfate (SDS), and 200 mM Tris-HCl (pH 6.8). The resulting cell lysates were then heated at 100°C for 5 min and protein concentrations were determined with the aid of the DC™ protein assay kit (Bio-Rad; Cat. No.: 5000111) following the manufacturer's recommendations. Next, 20-µg protein samples were separated by 6% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and were subsequently transferred onto 0.45-µm polyvinylidene difluoride (PVDF) membranes. Non-specific antibody binding was blocked by incubating the membranes in 5% (w/v) non-fat dry milk (Campina Elk; Cat. No.: 112349) dissolved in Tris-buffered saline (TBS; 50 mM Tris-HCl pH 7.6; 150 mM NaCl) supplemented with 0.1% (v/v) Tween 20 (Merck Millipore; Cat. No.: 8221840500) (TBST) at RT for at least 1 h. The blocked membranes were immediately incubated with the primary antibodies directed against *S. pyogenes* Cas9 (Abcam; Cat. No.: ab191468) or vinculin (Sigma-Aldrich; Cat. No.: V9131), each diluted 1:1000 in TBST containing 5% BSA. After overnight incubation at 4°C, the membranes were washed thrice with TBST before being exposed to a horseradish peroxidase (HRP)-conjugated secondary antibody directed against mouse IgG (Sigma-Aldrich; Cat. No.: NA931V) diluted 1:5000 in TBST containing 1% non-fat dry milk at RT for 2 h. Clarity™ Western ECL Substrate (Bio-Rad; Cat. No.: 1705060) was applied for signal detection by using the ChemiDoc Imaging System (Bio-Rad).

RT-qPCR analyses

Tracing CRISPRoff-induced *CD81* silencing at the mRNA level was done via RT-qPCR. In brief, total RNA was first extracted from CD81-negative and CD81-positive HEK293T cells by using the NucleoSpin RNA Kit (Macherey-Nagel; Cat. No.: 740955) following the manufacturer's recommendations. Afterwards, equal amounts of isolated RNA templates were applied for reverse transcription with the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific; Cat. No.: K1691). In brief, 1 µg of RNA was incubated with 0.5 µl of 100 µM random hexamer primers and 0.5 µl of 100 µM Oligo(dT)$_{18}$ primers in 12 µl reaction volumes at 65°C for 5 min followed by an 2-min incubation at 4°C. Subsequently, 1 µl of 20 U µl^{-1} RiboLock RNase Inhibitor, 1 µl of 200 U µl^{-1} RevertAid H Minus M-MuLV Reverse Transcriptase, 2 µl of 10 mM dNTP Mix and 4 µl of 5× Reaction Buffer, were directly added to each sample and the resulting mixtures were incubated at 25°C for 5 min followed by an 1-h incubation at 42°C. Afterwards, the reverse transcriptase was deactivated by heating the samples at 70°C for 5 min. Next, the resulting cDNA was subjected to qPCR by using the primers listed in **Supplementary Table S17** together with the iQ™ SYBR® Green Supermix (Bio-Rad; Cat. No.: L010171C). The qPCR signals were detected with the aid of a CFX Connect Real-Time PCR Detection System (Bio-Rad). The relative *CD81* mRNA expression was analyzed through the $2^{-\Delta\Delta Ct}$ method with *GAPDH* transcripts serving as internal controls for gene expression normalization. The qPCR cycling conditions and mixture components used for the analysis of *CD81* mRNA expression are specified in **Supplementary Tables S17** and **S18**, respectively.

ChIP-qPCR analyses

Establishing the acquisition of euchromatic and heterochromatic marks at *CD81* alleles was performed via ChIP-qPCR analyses as follows. Briefly, 2×10^7 cells were crosslinked with 1% formaldehyde for 10 min at RT and then immediately quenched with 1.25 M glycine (Thermo Fisher Scientific; Cat. No.:

120070050) for 5 min at RT. Next, the cells were washed twice with ice-cold PBS (pH 7.4) before being lysed in lysis buffer containing 5 mM piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES; pH 8.0; Sigma; Cat. No.: P6757), 85 mM KCl, 0.5% NP40 (Sigma; Cat. No.: 74385), and 1% proteinase inhibitor cocktail (Roche; Cat. No.: 11849300) for 10 min on ice. After a 5-min centrifugation at 510 ×g at 4°C, the sample supernatants were removed and the resulting nuclei portions were directly subjected to nuclei lysis buffer consisting of 50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS, and 1% proteinase inhibitor cocktail for 10 min on ice. Next, the chromatin was sheared by using a sonication device (Diagenode) set to optimized conditions predefined to obtain DNA fragments spanning the 200-bp to 700-bp range, i.e., 3 sec on, 6 sec off, 30% amplitude for 3 min. The sonicated samples were subsequently centrifuged at 17,949 ×g for 20 min at 4°C, and the resulting supernatants were collected and diluted 5-fold in immunoprecipitation (IP) dilution buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.25% sodium deoxycholate (Sigma; Cat. No.: D6750), and 1% proteinase inhibitor cocktail). A fraction of the cell lysis material (5% of total) corresponding to each sample was saved as input for qPCR normalization. The remaining cell lysis material of each sample was precleared with Protein A/G Sepharose beads (GE Healthcare; Cat. No.: 17-0963-03 or 17-0618-02) and salmon sperm DNA for at least 1 h at 4°C with agitation. Afterwards, immunoprecipitations were carried out by incubating the aforesaid cell lysis materials overnight at 4°C with ChIP-grade antibodies raised against H3K9me3 (10 µg; rabbit; Active Motif; Cat. No.: 39765), H3Ac (10 µg; rabbit; Active Motif; Cat. No.: 61637), RNA Pol II (2.5 µg; mouse; Active Motif; Cat. No.: 39097), IgG (10 µg; rabbit; Cell Signaling Technology; Cat. No.: 2729) or IgG (2.5 µg; mouse; Cell Signaling Technology; Cat. No.: 5415). In parallel, appropriate amounts of Protein A/G Sepharose beads were blocked overnight in 1% BSA. The next day, 100 µl of blocked beads were added to the cell lysis samples. After a 2-h incubation at 4°C with agitation, the beads were harvested and washed by using the following protocol: once with IP wash buffer #1 for 5 min at RT (20 mM Tris pH 8.0, 2 mM EDTA, 50 mM KCl, 1% Triton X-100 (Merck Millipore; Cat. No.: 1086031000), and 0.1% SDS), twice with high salt buffer for 5 min at RT (20 mM Tris pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, and 0.01% SDS), once with IP wash buffer #2 for 5 min at RT (10 mM Tris pH 8.0, 1 mM EDTA, 0.25 lithium chloride (J.T. Baker; Cat. No.: 0516), 1% NP40, and 1% sodium deoxycholate), and twice with TE buffer #1 for 5 min at RT (10 mM Tris pH 8.0, 1 mM EDTA). After these washing steps, the DNA in the various samples was eluted in elution buffer containing 50 mM sodium bicarbonate and 1% SDS at 65°C for 1 h and the subjected treated with 5 ml of 10 mg ml^{-1} RNase A (Thermo Fisher Scientific; Cat. No.: EN0531) overnight at 37°C, followed by a 5-h incubation with 2.5 ml of 20 mg ml^{-1} proteinase K (Thermo Fisher Scientific; Cat. No.: EO0491) at 55°C. Subsequently, the resulting DNA was precipitated by incubation overnight at -80°C in 500 ml of isopropanol supplemented with 2 ml of 20 mg ml^{-1} glycogen (Thermo Fisher Scientific; Cat. No.: R0551). The DNA pellets were harvested by centrifugation at 17,949 ×g for at least 30 min at 4°C, and were then washed once with 70% ethanol before being dissolved in 100 µl of TE buffer #2 consisting of 1 M Tris pH 8.0, 0.5 M EDTA and 20 mg ml^{-1} RNase A for 1 h at 37°C. Next, 1-µl samples of recovered purified DNA served as template for qPCR quantification by using the iQ™ SYBR® Green Supermix. The primers, cycling conditions and components of the qPCR mixtures are specified in **Supplementary Tables S17 and S18**. Finally, the signal outputs were detected with the aid of a CFX Connect Real-Time PCR Detection System (Bio-Rad).

Deep sequencing analysis

Amplicon deep sequencing was performed for assessing gene editing frequencies and outcomes resulting from the delivery of prime editing and base editing tools into human cells. The HEK.EGFP$^{TetO.KRAB}$ cells were exposed to the transfection mixtures indicated in **Supplementary Tables S11**-**S16**. At 17 days post-transfection, the frequencies of EGFP-negative cells and EBFP-positive cells were determined by flow cytometry as a live-cell readout for gene editing events. In parallel, genomic DNA was extracted with the DNeasy Blood & Tissue kit following the manufacturer's recommendations. Similarly, CD81-negative and CD81-positive HEK293T cells were subjected to prime editing or base editing complexes and, at 10 days post-transfection, genomic DNA was isolated for determining the gene-editing frequencies at the *CD81* locus. In brief, isolated DNA served as template in target-specific PCR mixtures containing Phusion High-Fidelity Polymerase (Thermo Fisher Scientific; Cat. No.: F-530L) and primers possessing adapter tag overhangs. After purification with AMPure XP beads (Beckman Coulter; Cat. No.: A63881), the resulting amplicons were subjected to PCR barcoding using Illumina tag-specific primer pairs possessing unique sequence identifier combinations. The cycling conditions, primer sequences and PCR mixture compositions are specified in **Supplementary Tables S19**-**S23**.

Prior to proceeding with amplicon sample quality control, the barcoded amplicons were purified with AMPure XP beads and their concentrations were measured by using the Qubit dsDNA HS assay kit (Thermo Fisher Scientific; Cat. No.: Q32854) together with a Qubit2.0 fluorometer. Afterwards, amplicon sample quality control was done by capillarity electrophoresis through a 2100 Bioanalyzer system (Agilent) with the Agilent DNA 1000 Kit. Finally, purified amplicons were pooled together in equal molar ratios and were then subjected to Illumina MiSeq deep sequencing for obtaining 50,000 paired-end reads. The paired-end MiSeq raw reads (R1 and R2 fastq files) were subjected to demultiplexing and then analysed with the aid of the CRISPResso2 software[63].

Statistical analyses

GraphPad Prism software (version 8.0.1) was used in the statistical analyses of datasets derived from at least three independent biological replicates. The statistical tests used and resulting significance outputs are, where relevant, indicated in the figures and respective legends.

DATA AVAILABILITY

All data assembled for and analysed in this study are included in the article and supplementary files. The libraries of next-generation sequencing reads are deposited at the NCBI Sequence Read Archive (SRA) database. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors thank all laboratory members for insightful discussions and members of Leiden Genome Technology Centre and Flow Cytometry Core Facility (Leiden University Medical Center, The Netherlands) for assistance with genomic sequencing and FACS services, respectively. Qian Wang held a Ph.D. fellowship from the China Scholarship Council–Leiden University Joint Scholarship Programme. *Conflict of interest statement*. None declared.

REFERENCES

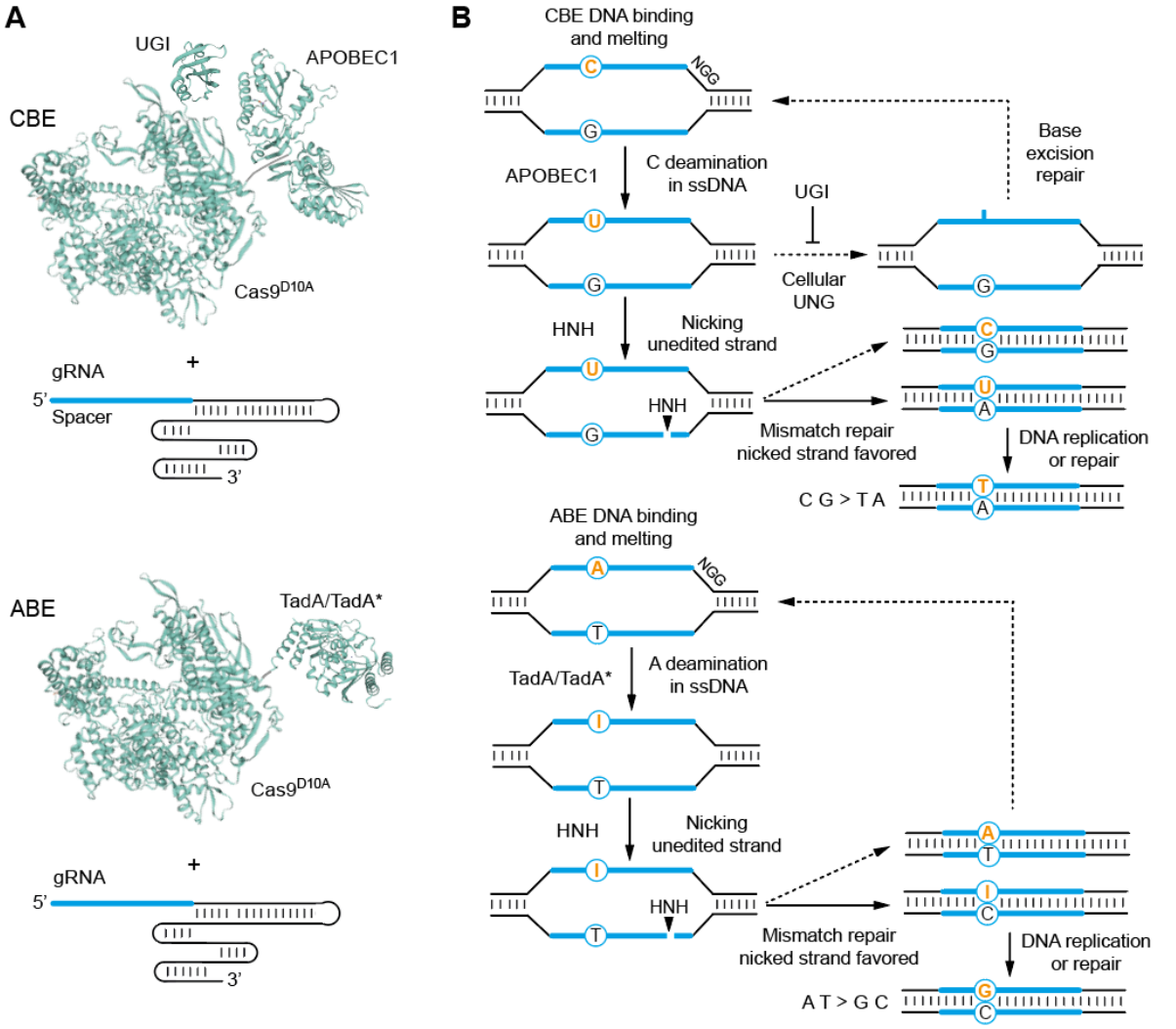
1. Chen, X., and Gonçalves, M.A.F.V. DNA, RNA, and protein tools for editing the genetic information in human cells. *iScience.* 2018; **6**: 247-263.
2. Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 2013; **41**: 9584–9592.
3. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013; **31**: 822–826.
4. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 2013; **31**: 827–832.
5. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu D.R. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Bioatechnol.* 2013; **31**: 839–843.
6. Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt F.W. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 2015; **33**: 179–186.
7. Chen, X., Tasca, F., Wang, Q., Liu, J., Janssen, J.M., Brescia, M.D., Bellin, M., Szuhai, K., Kenrick, J., Frock, R.L., et al. Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking. *Nucleic Acids Res.* 2020; **48**: 974-995.
8. Wang, Q., Liu, J., Janssen, J.M., Le Bouteiller, M., Frock, R.L., and Gonçalves, M.A.F.V. Precise and broad scope genome editing based on high-specificity Cas9 nickases. *Nucleic Acids Res.* 2021; **49**: 1173–1198.
9. Liu, M., Zhang, W., Xin, C., Yin, J., Shang, Y., Ai, C., Li, J., Meng, F.L., and Hu, J. Global detection of DNA repair outcomes induced by CRISPR-Cas9. *Nucleic Acids Res.* 2021; **49**: 8732-8742.
10. Turchiano, G., Andrieux, G., Klermund, J., Blattner, G., Pennucci, V., El Gaz, M., Monaco, G., Poddar, S., Mussolino, C., Cornu, T.I., et al. Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell.* 2021; **28**: 1136–1147.

11. Kosicki, M., Tomberg, K., and Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 2018; **36**: 765–771.
12. Boutin, J., Rosier, J., Cappellen, D., Prat, F., Toutain, J., Pennamen, P., Bouron, J., Rooryck, C., Merlio, J.P., Lamrissi-Garcia, I., et al. CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. *Nat. Commun.* 2021; **12**: 4922.
13. Leibowitz, M.L., Papathanasiou, S., Doerfler, P.A., Blaine, L.J., Sun, L., Yao, Y., Zhang, C.Z., Weiss, M.J., and Pellman, D. Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.* 2021; **53**:895-905.
14. Nahmad, A.D., Reuveni, E., Goldschmidt, E., Tenne, T., Liberman, M., Horovitz-Fried, M., Khosravi, R., Kobo, H., Reinstein, E., Madi, A., et al. Frequent aneuploidy in primary human T cells after CRISPR-Cas9 cleavage. *Nat. Biotechnol.* 2020; **40**: 1807-1813.
15. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects *Nature*. 2016; **529**: 490-495.
16. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016; **351**: 84-88.
17. Chen, X., Janssen, J.M., Liu, J., Maggio, I., Jong, A.E.J., Mikkers, H.M.M., and Gonçalves, M.A.F.V. In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun.* 2017; **8**: 657.
18. Nakajima, K., Zhou, Y., Tomita, A., Hirade, Y., Gurumurthy, C.B., and Nakada, S. Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. *Genome Res.* 2018; **28**: 223–230.
19. Hyodo, T., Rahman, M.L., Karnan, S., Ito, T., Toyoda, A., Ota, A., Wahiduzzaman, M., Tsuzuki, S., Okada, Y., Hosokawa, Y., et al. Tandem paired nicking promotes precise genome editing with scarce interference by p53. *Cell Rep.* 2020; **30**: 1195–1207.
20. Bollen, Y., Hageman, J.H., van Leenen, P., Derks, L.L.M., Ponsioen, B., Buissant des Amorie, J.R., Verlaan-Klink, I., van den Bos, M., Terstappen, L.W.M.M., van Boxtel, R., et al. Efficient and error-free fluorescent gene tagging in human organoids without double-strand DNA cleavage. *PLoS Biol.* 2022; **20**:e3001527.
21. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016; **533**: 420-424.
22. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. 2016; **353**: p. aaf8729.
23. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*. 2017; **551**: 464-471.
24. Jeong, Y.K., Song, B., and Bae, S. Current Status and Challenges of DNA Base Editing Tools. *Mol. Ther.* 2020; **28**:1938-1952.
25. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 2019; **576**: 149–157.
26. Chen, P.J., Hussmann, J.A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.F., Chen, C., Nelson, J.W., Newby, G.A., Sahin, M., et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*. 2021; **184**: 5635-5652.e29.
27. Wang, Q., Liu, J., Janssen, J.M., Tasca, F., Mei, H., and Gonçalves, M.A.F.V. Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. *Nucleic Acids Res.* 2021; **49**:11986-12001.
28. Ferreira da Silva, J., Oliveira, G.P., Arasa-Verge, E.A., Kagiou, C., Moretton, A., Timelthaler, G., Jiricny, J., and Loizou, J.I. Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat. Commun.* 2020; **13**: 760.
29. Chen, X., Rinsma, M., Janssen, J.M., Liu, J., Maggio, I., and Gonçalves, M.A.F.V. Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res*. 2016; **44**: 6482-92.
30. Dear, R.M., Cutts, J.P., Brafman, D.A., and Haynes, K.A. The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. *ACS Synth. Biol.* 2017; **6**:428-438.

31. Chen, X., Liu, J., Janssen, J.M., and Gonçalves, M.A.F.V. The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Mol. Ther. Nucleic Acids*. 2017; **8**: 558-563.
32. Jensen, K.T., Fløe, L., Petersen, T.S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett.* 2017; **591**:1892-1901.
33. Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome Biol,* 2003; **4**: 231.
34. Szulc, J., Wiznerowicz, M., Sauvain, M.O., Trono, D., and Aebischer P. (2006) A versatile tool for conditional gene expression and knockdown. *Nat. Methods*. 2006; **3**: 109-116.
35. Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet*. 2010; **6**, e1000869.
36. Nuñez, J.K., Chen, J., Pommier, G.C., Cogan, J.Z., Replogle, J.M., Adriaens, C., Ramadoss, G.N., Shi, Q., Hung, K.L., Samelson, A.J., et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell*. 2021; **184**: 2503-2519.
37. Jiang, G., Wang, J., Zhao, D., Chen, X., Pu, S., Zhang, C., Li, J., Li, Y., Yang, J., Li, S., et al. Molecular mechanism of the cytosine CRISPR base editing process and the roles of translesion DNA polymerases. *ACS Synth. Biol.* 2021; **10**: 3353-3358.
38. Park, S.J., Jeong, T.Y., Shin, S.K., Yoon, D.E., Lim, S.Y., Kim, S.P., Choi, J., Lee, H., Hong, J.I., Ahn, J., et al. Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome Biol.* 2021; **22**: 170.
39. Chen, R., Cao, Y., Liu, Y., Zhao, D., Li, J., Cheng, Z., Bi, C., and Zhang, X. Enhancement of a prime editing system via optimal recruitment of the pioneer transcription factor P65. *Nat. Commun.* 2023; **14**: 257.
40. Nelson, J.W., Randolph, P.B., Shen, S.P., Everette, K.A., Chen, P.J., Anzalone, A.V., An, M., Newby, G.A., Chen, J.C., Hsu, A. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* 2022; **40**, 402-410.
41. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 2020; **38**: 883-891.
42. Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., and Liu, D.R. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol*. 2018; *36*: 843-846.
43. Yang, C., Dong, X., Ma, Z., Li, B., Bi, C., and Zhang, X. Pioneer Factor Improves CRISPR-Based C-To-G and C-To-T Base Editing. *Adv. Sci. (Weinh)*. 2022; **9**: e2202957.
44. Yang, C., Ma, Z., Wang, K., Dong, X., Huang, M., Li, Y., Zhu, X., Li, J., Cheng, Z., Bi, C., and Zhang, X. HMGN1 enhances CRISPR-directed dual-function A-to-G and C-to-G base editing. *Nat. Commun.* 2023; **14**: 2430.
45. Shin, H.R., See, J.E., Kweon, J., Kim, H.S., Sung, G.J., Park, S., Jang, A.H., Jang, G., Choi, K.C., Kim, I., Kim, J.S., and Kim, Y. Small-molecule inhibitors of histone deacetylase improve CRISPR-based adenine base editing. *Nucleic Acids Res*. 2021; **49**: 2390-2399.
46. Liu, N., Zhou, L., Lin, G., Hu, Y., Jiao, Y., Wang, Y., Liu, J., Yang, S., and Yao, S. HDAC inhibitors improve CRISPR-Cas9 mediated prime editing and base editing. *Mol. Ther. Nucleic Acids.* 2022; **29**: 36-46.
47. Roos, W.P., and Krumm, A. The multifaceted influence of histone deacetylases on DNA damage signalling and DNA repair. *Nucleic Acids Res*. 2016; **44**: 10017-10030.
48. Winter, J., Luu, A., Gapinske, M., Manandhar, S., Shirguppe, S., Woods, W.S., Song, J.S., and Perez-Pinera, P. Targeted exon skipping with AAV-mediated split adenine base editors. *Cell Discov*. 2019; **5**: 41.
49. Song, M., Kim, H.K., Lee, S., Kim, Y., Seo, S.Y., Park, J., Choi, J.W., Jang, H., Shin, J.H., Min, S., et al. Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nat. Biotechnol*. 2020; **38**: 1037-1043.
50. Lee, H.W., Dominy, B.N., and Cao, W. New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *J. Biol. Chem*. 2011; **286**: 31282-7.
51. Beard, B.C., Wilson, S.H., and Smerdon, M.J. Suppressed catalytic activity of base excision repair enzymes on rotationally positioned uracil in nucleosomes. *Proc. Natl. Acad. Sci. USA.* 2003; **100**: 7465–7470.
52. Hinz, J.M., Rodriguez, Y., and Smerdon, M.J. Rotational dynamics of DNA on the nucleosome surface markedly impact accessibility to a DNA repair enzyme. *Proc. Natl. Acad. Sci. USA.* 2010; **107**: 4646–4651.
53. Rodriguez, Y., and Smerdon, M.J. The structural location of DNA lesions in nucleosome core particles determines accessibility by base excision repair enzymes. *J. Biol. Chem*. 2013; **288**: 13863-13875.
54. Hinz, J.M. Impact of abasic site orientation within nucleosomes on human APE1 endonuclease activity. *Mutat Res.* 2014; **766–767**: 19–24.

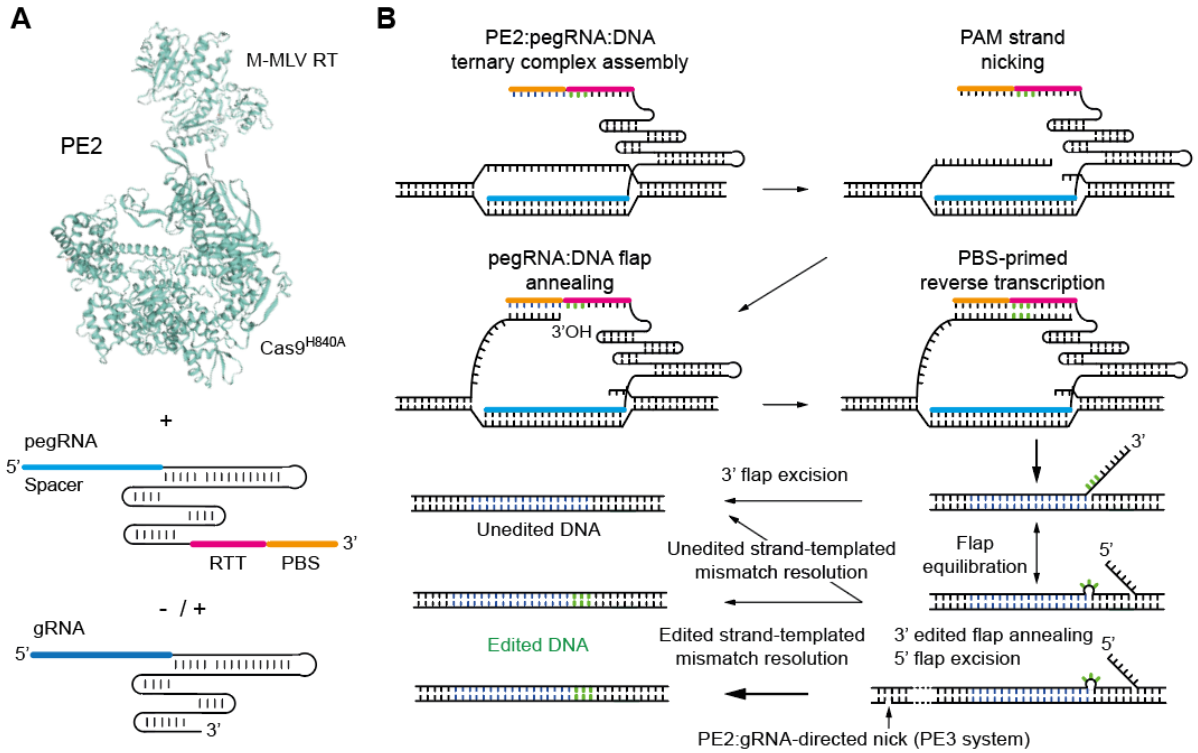
55. Menoni, H., Gasparutto, D., Hamiche, A., Cadet, J., Dimitrov, S., Bouvet, P., and Angelov, D. ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A Bbd nucleosomes. *Mol. Cell Biol*. 2007; **27**: 5949–5956.
56. Amouroux, R., Campalans, A., Epe, B., and Radicella, J.P. Oxidative stress triggers the preferential assembly of base excision repair complexes on open chromatin regions. *Nucleic Acids Res.* 2010; **38**: 2878-2890.
57. Montaldo, N.P., Bordin, D.L., Brambilla, A., Rösinger, M., Fordyce Martin, S.L., Bjørås, K.Ø., Bradamante, S., Aas, P.A., Furrer, A., Olsen, L.C., et al. Alkyladenine DNA glycosylase associates with transcription elongation to coordinate DNA repair with gene expression. *Nat. Commun.* 2019; **10**: 5460.
58. Kingwell, K. Base editors hit the clinic. *Nat. Rev. Drug Discov*. 2022; **21**: 545-547.
59. Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J., and Scharenberg, A.M.. Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods*. 2011; **8**: 671-676.
60. Maggio, I., Zittersteijn, H.A., Wang, Q., Liu, J., Janssen, J.M., Ojeda, I.T., van der Maarel, S.M., Lankester, A.C., Hoeben, R.C., and Gonçalves, M.A.F.V. Integrating gene delivery and gene-editing technologies by adenoviral vector transfer of optimized CRISPR-Cas9 components. *Gene Ther*. 2020; **27**: 209-225.
61. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. RNA-guided human genome engineering via Cas9. *Science*. 2013; **339**: 823-826.
62. Zafra, M.P., Schatoff, E.M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A.Y., Simon, A., Han, T., Goswami, S., Montgomery, E., et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol*. 2018; **36**: 888-893.
63. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol*. 2019; **37**: 224-226.

SUPPLEMENTARY FIGURES

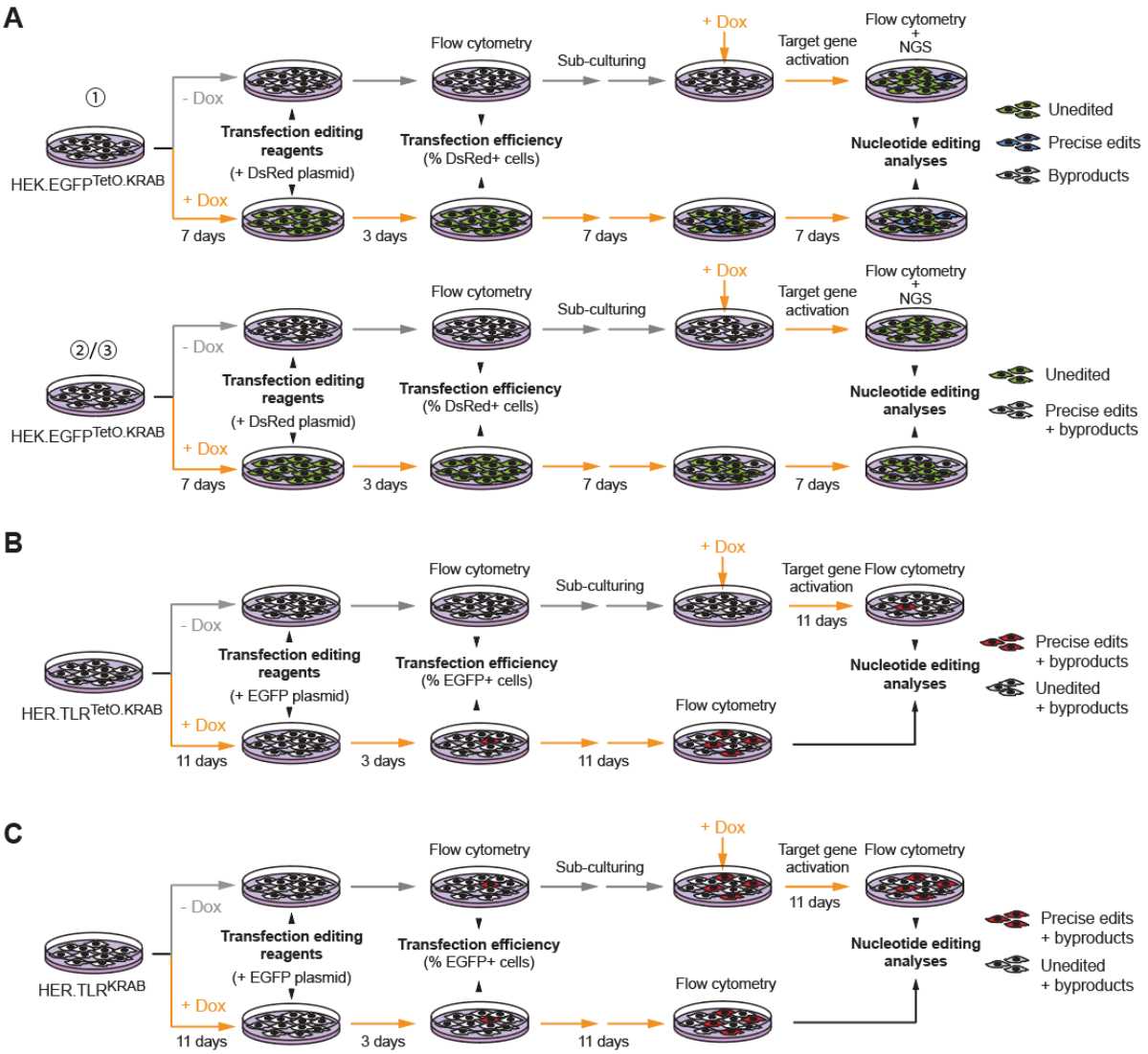


Supplementary Figure S1. Base editing systems. (A) Base editing elements. Cytosine base editors (CBEs) yield C•G to T•A substitutions and are formed by the fusion of a $Cas9^{D10A}$ nickase to a cytosine deaminase, often APOBEC1, and an uracil DNA

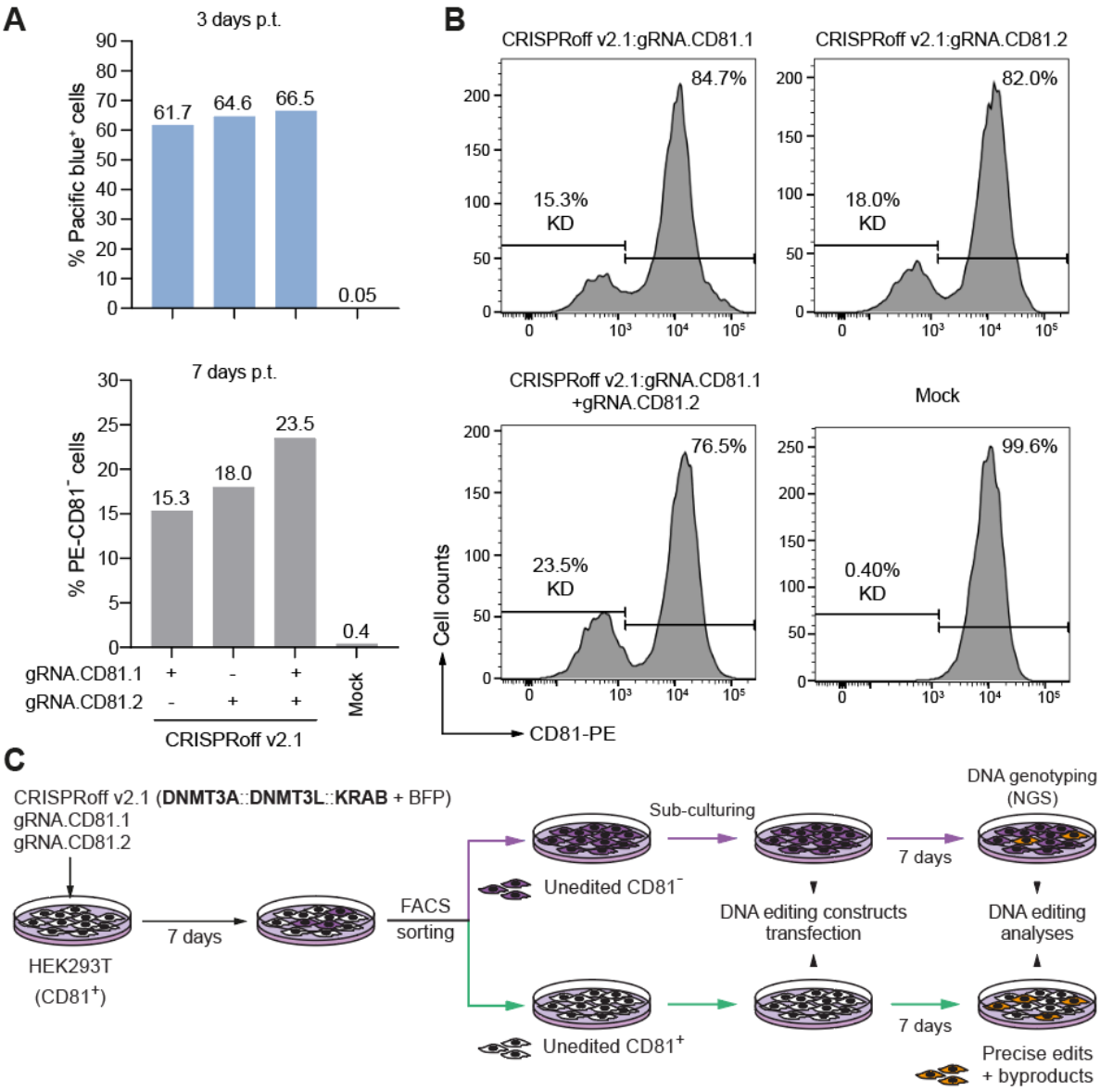
glycosylase inhibitor (UGI). Adenine base editors (ABEs) yield A•T to G•C substitutions and are formed by the fusion of a $Cas9^{D10A}$ nickase to an engineered *Escherichia coli* adenine deaminase consisting of a TadA-TadA* heterodimer. In *E. coli* TadA forms a homodimer. One monomer converts adenine to inosine (I) in tRNA and the other assists in substrate binding. In ABEs, the non-catalytic wild-type TadA aids instead an evolved TadA* monomer in catalysing adenine deamination in single-stranded DNA instead of RNA. Both base editor types are addressed to target sequences through a regular gRNA. **(B)** Cytosine and adenine base editing *modus operandi*. CBE:gRNA binding to the target sequence forms an R loop exposing a region of single-stranded DNA. Cs in this single-stranded protospacer bubble become targets for the cytosine deaminase and convert into Us, especially those found in the so-called "activity window" whose position and length depends on the specific base editor architecture. The counterproductive activity of cellular uracil *N*-glycosylases (UNGs) involved in base excision repair of U•G intermediates is inhibited through the UGI moiety. Subsequently, nicking of the strand containing the original G induces cellular mismatch repair of this unedited stand resulting in G-to-A replacement. Finally, upon DNA repair or replication, conversion of the initial C•G into T•A, is completed. ABE:gRNA complexes trigger a series of DNA processing steps similar to those induced by CBE:gRNA complexes except that, upon R loop formation, As exposed in the single-strand DNA bubbles are deaminated by TadA-TadA* to I intermediates. These intermediates are subsequently converted into Cs through DNA repair or replication.



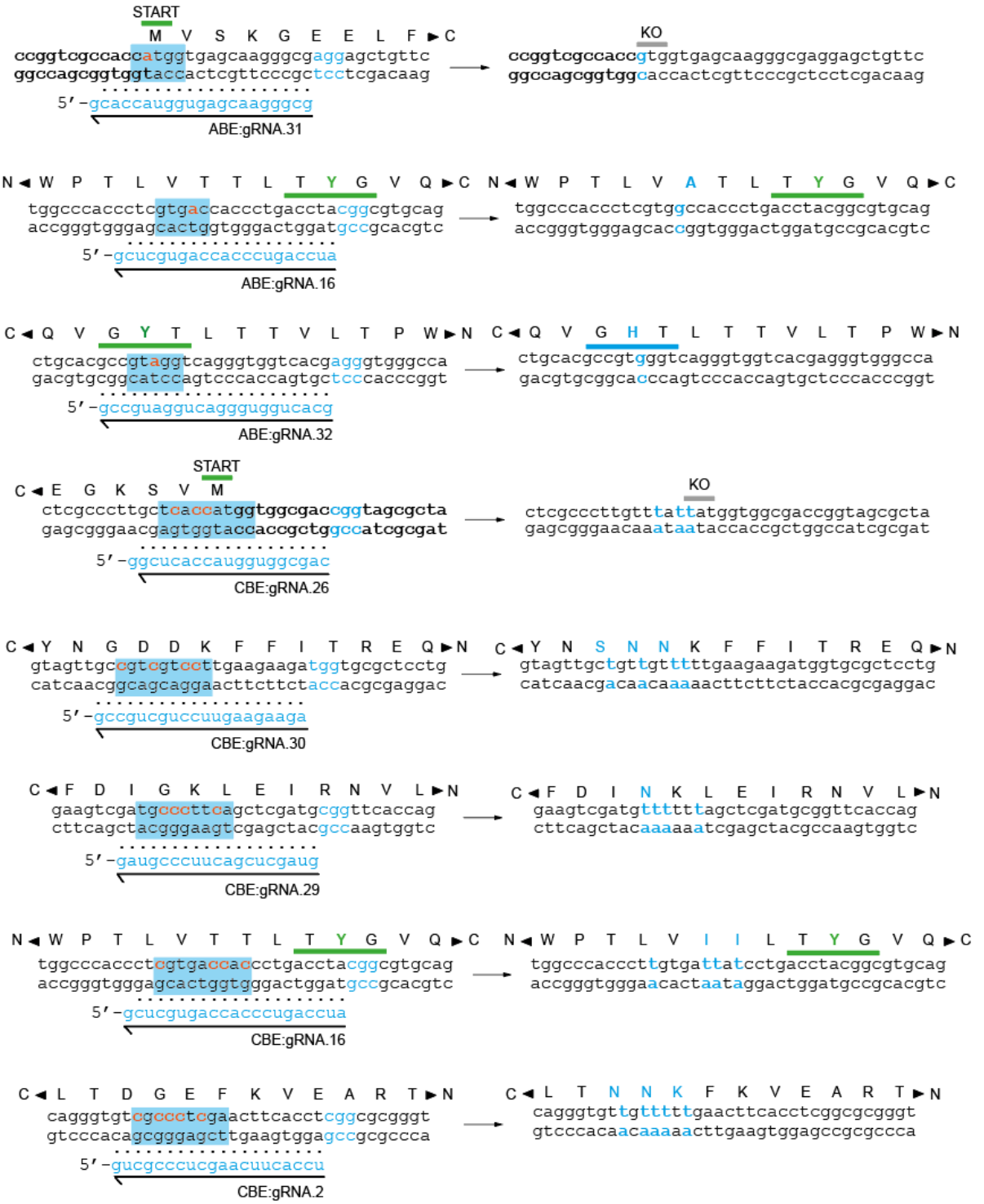
Supplementary Figure S2. Prime editing systems. (A) Prime editing elements. Prime editors are ribonucleoprotein complexes formed by a PE2 protein and a prime editor gRNA (pegRNA). The former element is a fusion product between the $Cas9^{H840A}$ nickase and an engineered Moloney murine leukemia virus reverse transcriptase (M-MLV RT); the latter element is a gRNA extended at the 3' end with reverse transcriptase template (RTT) and primer binding site (PBS) sequences. PE2:pegRNA complexes form the PE2 system. The addition of an auxiliary gRNA forms the PE3 system that can enhance prime editing activities, albeit with increases risks for bystander DSB-induced indel formation resulting from coordinated nicking of both DNA strands. **(B)** Prime editing *modus operandi*. The PE2 protein is addressed to a target sequence through the spacer portion of a pegRNA (PE2 protein not drawn for clarity). At the target site, site-specific nicking releases a single-stranded DNA flap that, after hybridizing to the complementary PBS, provides a free 3'-OH group for M-MLV RT-mediated RNA-dependent DNA polymerization (reverse transcription) over the edit-encoding RTT sequence. Hybridization of the *de novo* synthesized cDNA strand to complementary target DNA and excision of the resulting 5' flap forms heteroduplexes containing edited and unedited strands whose mismatches are further processed to yield edited and unedited homoduplexes. A gRNA directing non-edited strand nicking working in concert with a PE2:pegRNA complex (PE3 system) can enhance the accumulation of the desired edited homoduplexes, presumably through guiding DNA mismatch repair.



Supplementary Figure S3. Detailed schematics and time courses of DNA editing experiments. The tTR-KRAB-expressing reporter cells HEK.EGFP$^{TetO.KRAB}$ **(A)** and HER.TLR$^{TetO.KRAB}$ **(B)** were used for tracking and quantifying DNA editing outcomes induced by PE2, PE3, CBE and ABE reagents at euchromatic versus heterochromatic nucleotide sequences. The TetO-negative and tTR-KRAB-expressing reporter cells HER.TLR^{KRAB} **(C)** provided for negative controls. The HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$ systems permit assessing DNA editing settings resulting in both gain-of-function and loss-of-functions phenotypes as indicated. The initial higher-order chromatin conformation of target sites in both model alleles is controlled through Dox-dependent regulation of tTR-KRAB binding. HEK.EGFP$^{TetO.KRAB}$ and HER.TLR$^{TetO.KRAB}$ cells with target sites in a heterochromatic (–Dox) or euchromatic (+Dox) state, are transiently transfected with different gene editing constructs. DsRed and EGFP expression plasmids included in the transfection mixtures permit determining transient transfection efficiencies. After the completion of the various nucleotide editing processes in each of the two parallel experimental settings (i.e., –Dox and +Dox), target gene expression is activated allowing quantifying the frequencies of precise and bystander gene editing events flow cytometry and next generation sequencing.

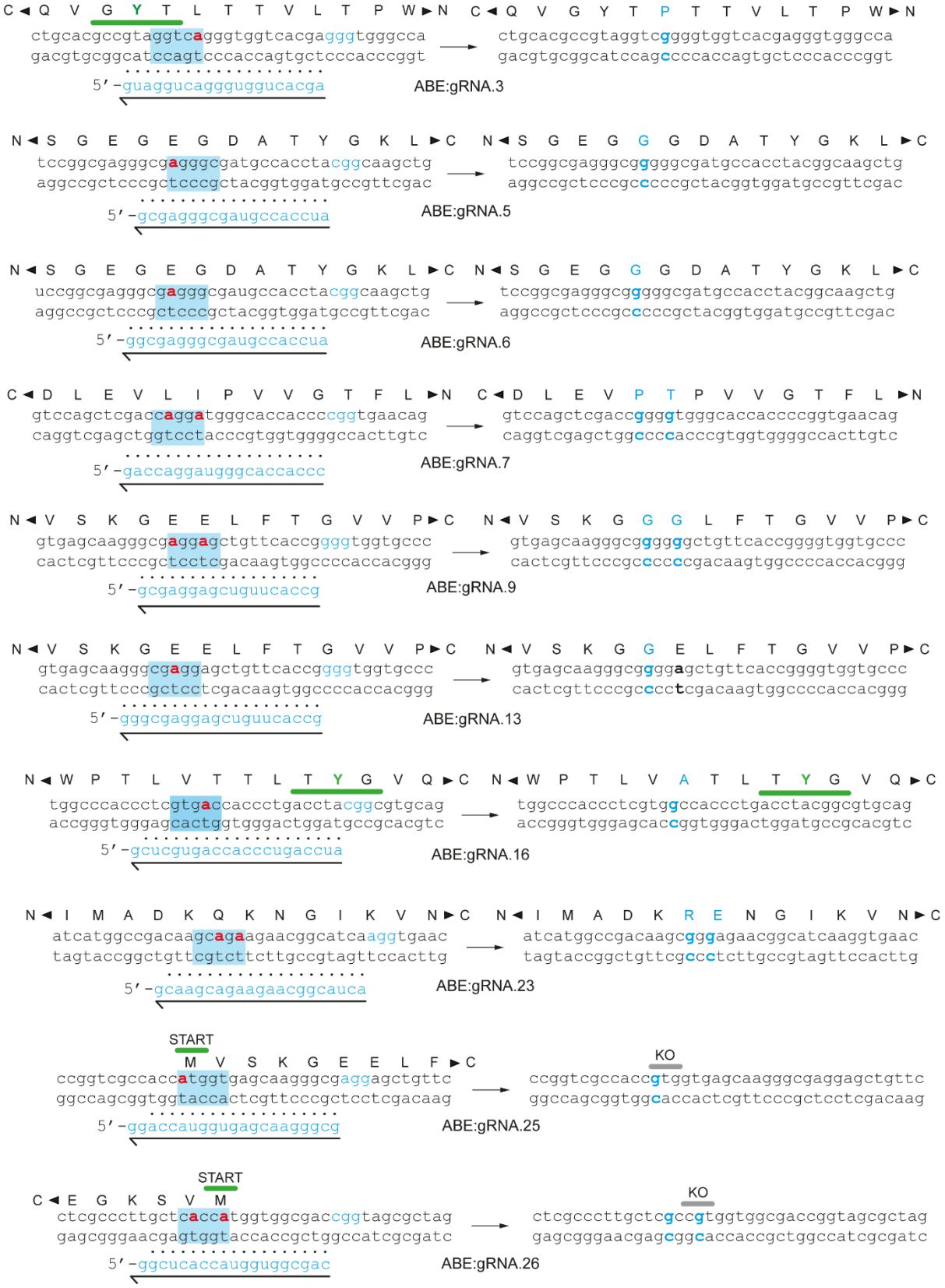


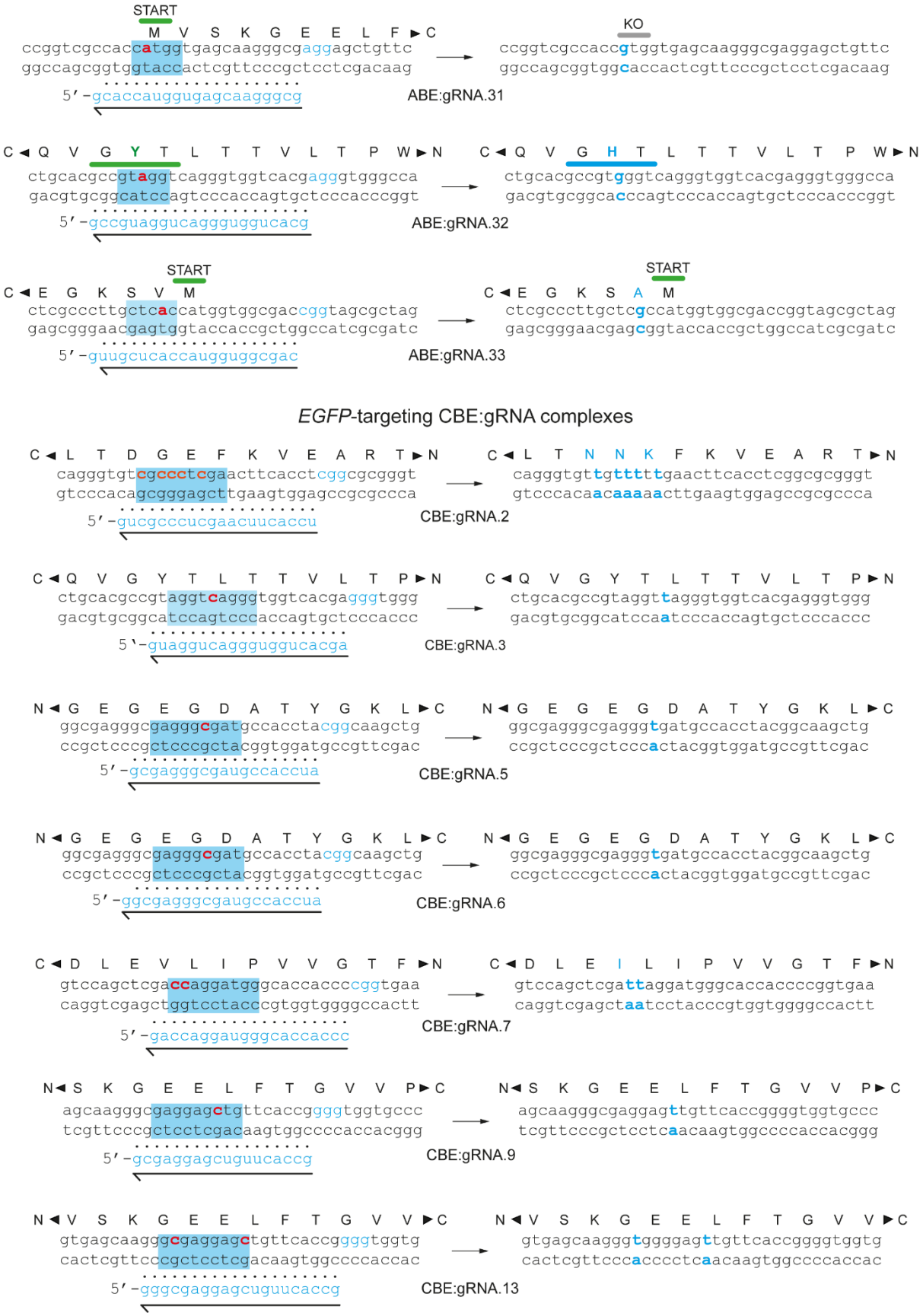
Supplementary Figure S4. Experimental set-up to investigate prime editing and base editing at endogenous sequences in open and close chromatin. (**A**) Epigenetic silencing of *CD81* alleles. HEK293T cells transfected with the indicated combinations of CRISPRoff and gRNA constructs were analysed by flow cytometry for BFP and CD81 expression at 3 and 7 days post-transfection, respectively. The CRISPRoff construct encodes the live-cell reporter BFP and a covalent protein assembly consisting of the DNA methyltransferases DNMT3A and DNMT3L fused to a chromatin remodelling KRAB domain. (**B**) Flow cytometry histograms corresponding to the data depicted in the bottom graph of panel A. (**C**) Diagram and time course of *CD81* gene editing experiments. HEK293T cell populations with *CD81* sequences in epigenetically native and silenced states generated through CRISPRoff transfection and fluorescence activated cell sorting (FACS) of $CD81^-$ and $CD81^+$ cell fractions.

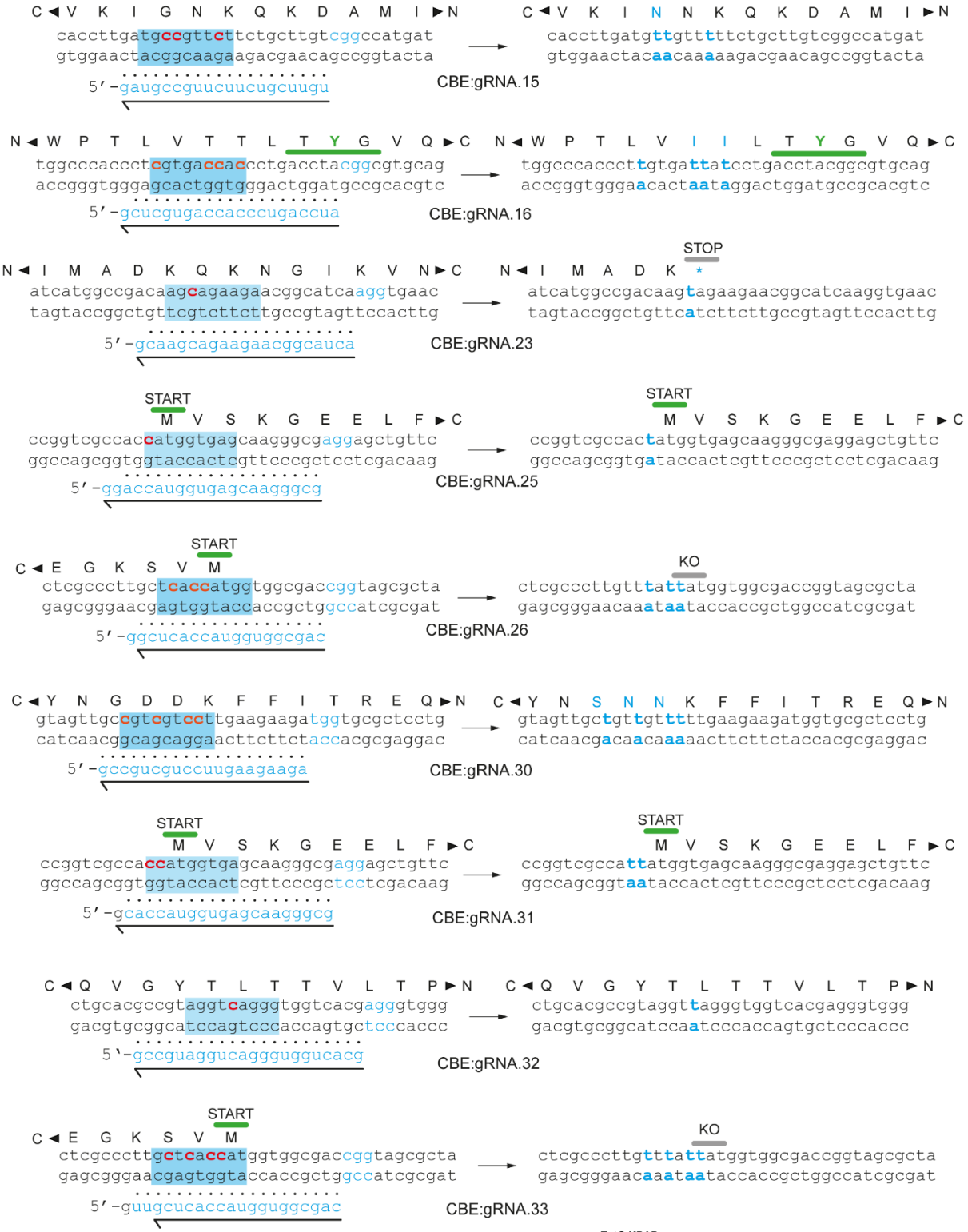


Supplementary Figure S5. Base editing complexes tested in HEK.EGFP$^{TetO.KRAB}$ system. Spacer sequences of gRNAs (horizontal arrows) used in base editing experiments (**Figure 6**) are drawn in relation to their target sites formed by protospacer and NGG PAM sequences. Substrate nucleotides and product base pairs predicted to result from base editing reactions, are highlighted in red and cyan lettering, respectively. Expected base editing windows of and amino acid changes induced by base editors are marked in cyan. EGFP and EBFP fluorophore sequences are underlined in green and cyan, respectively.

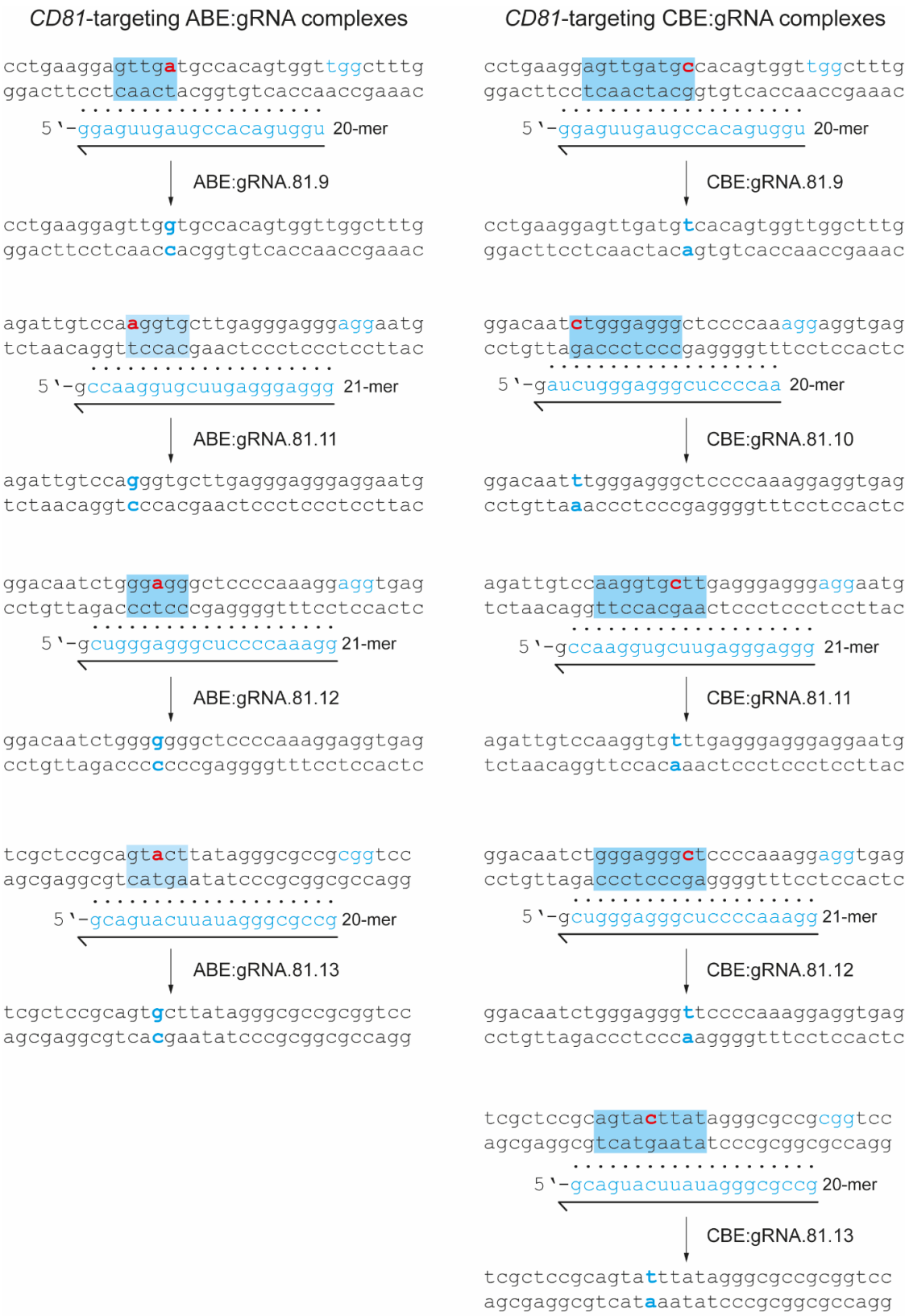
EGFP-targeting ABE:gRNA complexes



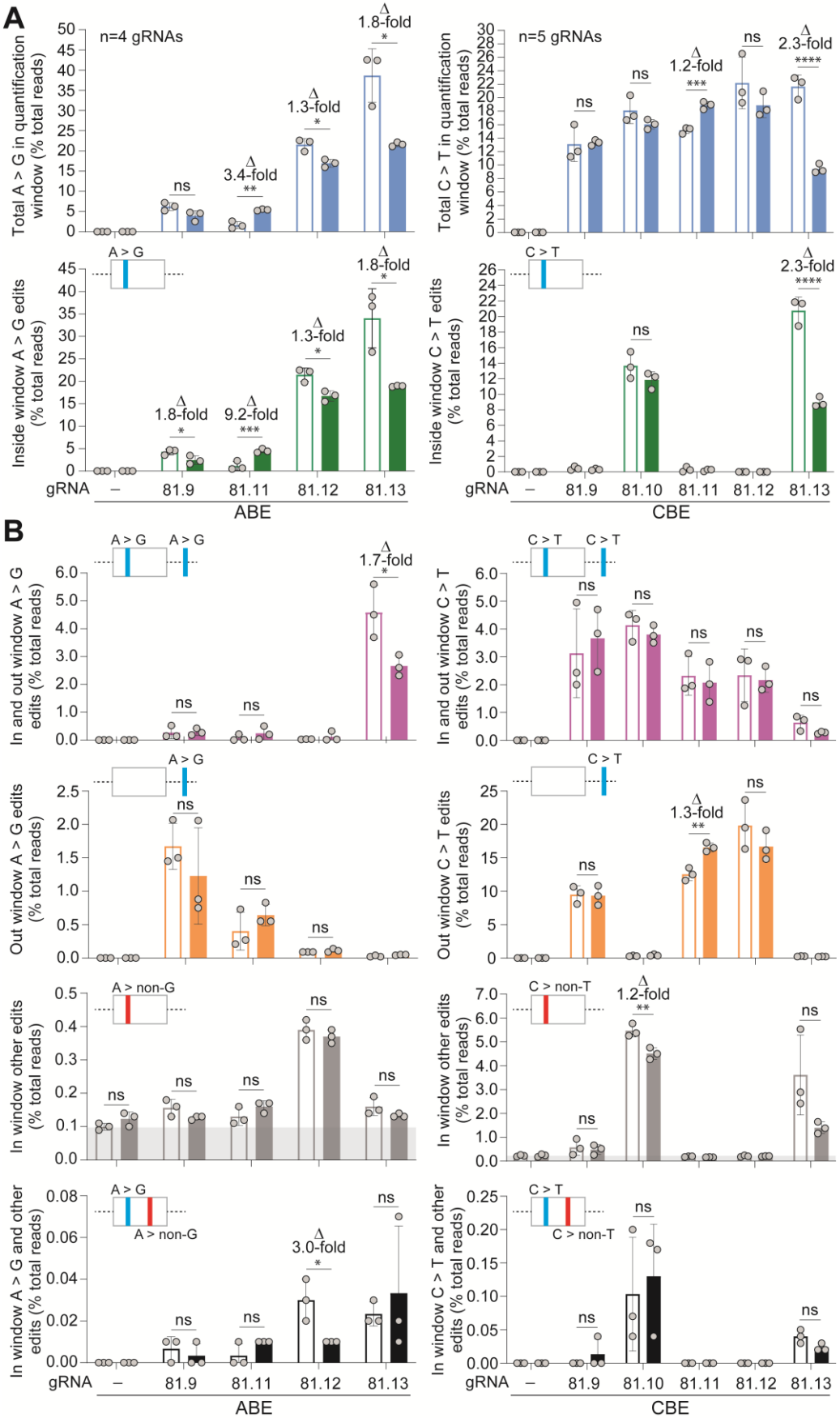

START
KO
ABE:gRNA.31
ABE:gRNA.32
START
START
ABE:gRNA.33
EGFP-targeting CBE:gRNA complexes
CBE:gRNA.2
CBE:gRNA.3
CBE:gRNA.5
CBE:gRNA.6
CBE:gRNA.7
CBE:gRNA.9
CBE:gRNA.13



Supplementary Figure S6. Base editing complexes tested in HEK.EGFP$^{TetO.KRAB}$ system. Spacer sequences of gRNAs (horizontal arrows) used in base editing experiments (**Figure 7**) are drawn in relation to their target sites formed by protospacer and NGG PAM sequences. Substrate nucleotides and product base pairs predicted to result from base editing reactions, are highlighted in red and cyan lettering, respectively. Expected base editing windows of and amino acid changes induced by base editors are marked in cyan. EGFP and EBFP fluorophore sequences are underlined in green and cyan, respectively.

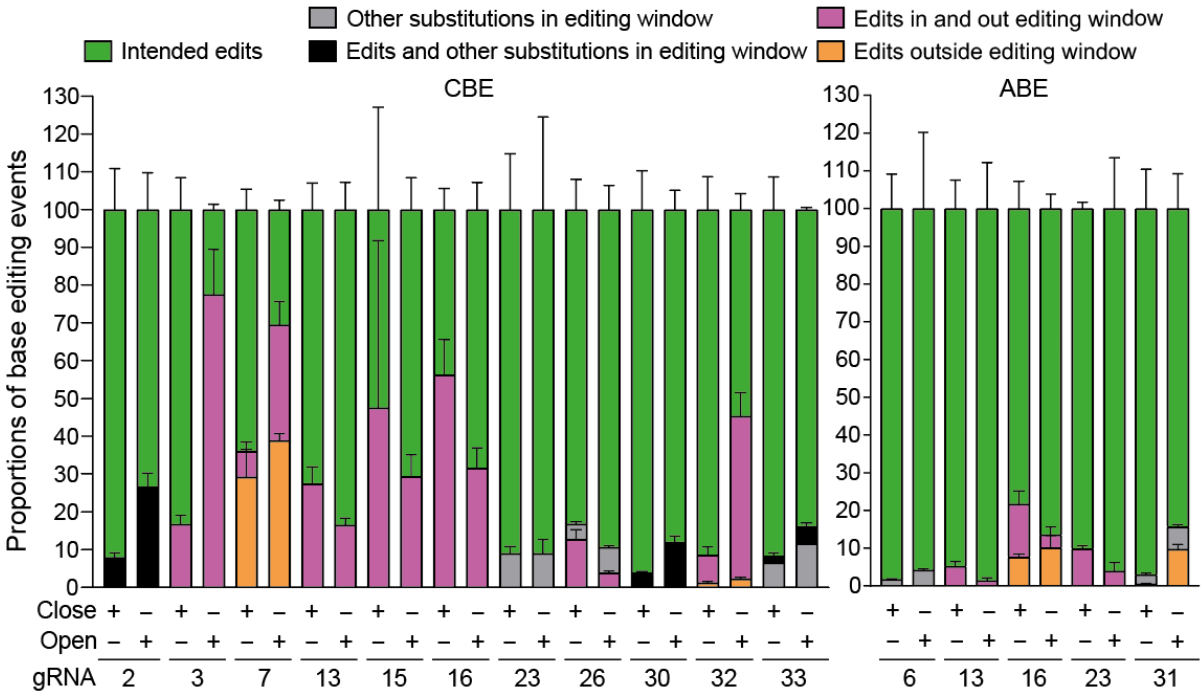


Supplementary Figure S7. Base editing complexes tested in heritable CD81^{+}/CD81^{-} HEK293T cell system. Spacer sequences of gRNAs (horizontal arrows) used in base editing experiments are drawn in relation to their target sites formed by protospacer and NGG PAM sequences. Substrate nucleotides and product base pairs predicted to result from base editing reactions, are highlighted in red and cyan lettering, respectively. Expected base editing windows of and amino acid changes induced by base editors are marked in cyan.

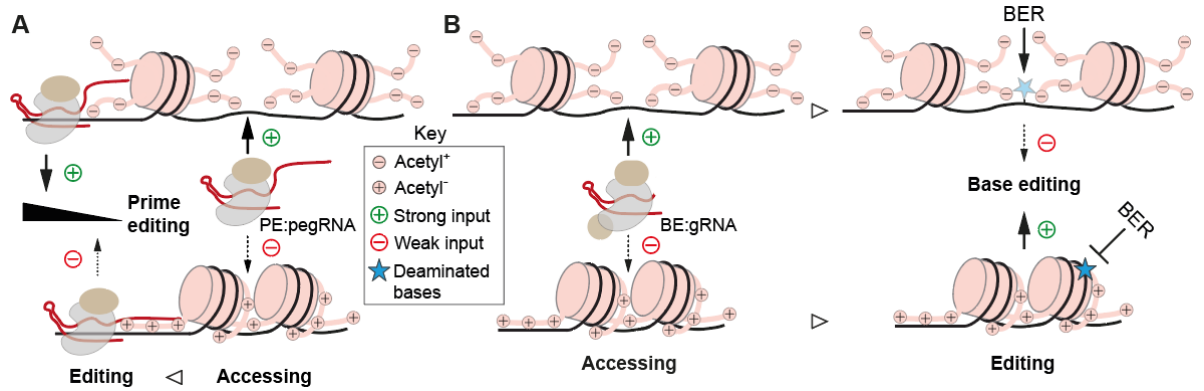


Supplementary Figure S8. Base editing at open and heritable closed chromatin. (**A**) Quantification of base edits at open and closed chromatin. HEK293T cells with active or epigenetically silenced *CD81* loci, were treated with the indicated ABE:gRNA and CBE:gRNA complexes designed for installing A > G and C > T substitutions, respectively, within their respective editing windows.

Total base edits within the quantification window and intended base edits within each editing window are plotted (top and bottom graphs, respectively). (**B**) Quantification of base-editing byproducts at open and closed chromatin. Base-editing byproducts corresponding to base edits inside and outside editing windows or only outside these windows are plotted as magenta and orange bar graphs, respectively. Base-editing byproducts corresponding to unintended substitutions and composite edits consisting of intended and unintended substitutions are plotted as grey and black bar graphs. Base editing events were measured through deep sequencing analyses (50,000 paired-end reads). Significances were determined via two-tailed Student's *t* tests with bars and error bars corresponding to mean ± s.d., respectively (n=3 biological replicates). $P > 0.05$ considered non-significant (ns).



Supplementary Figure S9. Proportions of base editing events generated by CBE and ABE complexes. Relative frequencies of the different types of base editing events resulting from CBE and ABE complexes leading to significant differences in the amounts of byproducts at open versus closed chromatin (plotted from Figure 7 datasets).



Supplementary Figure S10. Working models for the differential impact of chromatin organization on DSB-free gene editing tools. (**A**) Prime editing in open versus closed chromatin. PE:gRNA complexes access less frequently DNA in heterochromatin than euchromatin. Upon accessing heterochromatic target sites the 3'-ended extended portions of pegRNAs (i.e., PBS and RT templates) of prime editors locally interact with positively charged histone tails resulting in their trapping and consequent prime editing inhibition. Conversely, upon accessing euchromatic target sites these extended portions of pegRNAs are less likely to interact with acetylated negatively charged histone tails, resulting in their availability for DNA flap hybridization and reverse transcription. (**B**) base editing in open versus closed chromatin. As PE:gRNA complexes, BE:gRNA complexes access less frequently DNA in heterochromatin than in euchromatin. Upon accessing heterochromatic target sites, BE:gRNA complexes induce *in situ* deamination (stars). When compared to deaminated bases in euchromatin (light cyan star), deaminated bases in heterochromatin (dark cyan star) are more protected from reversion through base excision repair (BER) and, as a result, are more likely to remain a substrate for the downstream base editing processes. The postulated interactions (inputs) involving CRISPR complex accessibility to alternate chromatin conformations and local favouring or disfavouring of reverse transcriptase and deaminase activities results in a net negative or positive DNA editing output. Representative relative outputs of the herein investigated DSB-free DNA editing platforms in terms of chromatin impact indexes are as follows: PE:pegRNA > BE:gRNA.

SUPPLEMENTARY TABLES

Supplementary Table S1. Oligonucleotides used for gRNA assembly.

Plasmid name	Code	Oligonucleotide sequence (5' → 3')
AM51_pU6.gRNA$^{I\text{-}SceI.1}$	#25	ACCGGTGAGCTCTTATTTGCGTAGCTAGCTGAC
	#26	AAACGTCAGCTAGCTACGCAAATAAGAGCTCAC
BF23_pU6.opt-gRNA$^{eGFP.2}$ and AV59_pU6.gRNA$^{eGFP.2}$	#161	ACCGTCGCCCTCGAACTTCACCT
	#162	AAACAGGTGAAGTTCGAGGGCGA
BA21_pU6.opt-gRNA$^{eGFP.3}$	#163	ACCGTAGGTCAGGGTGGTCACGA
	#164	AAACTCGTGACCACCCTGACCTA
AW18_pU6.opt-gRNA$^{eGFP.5}$	#171	ACCGCGAGGGCGATGCCACCTA
	#172	AAACTAGGTGGCATCGCCCTCG
AW20_pU6.gRNA$^{eGFP.6}$	#173	ACCGGCGAGGGCGATGCCACCTA
	#174	AAACTAGGTGGCATCGCCCTCGC
AW22_pU6.gRNA$^{eGFP.7}$	#175	ACCGACCAGGATGGGCACCACCC
	#176	AAACGGGTGGTGCCCATCCTGGT
AW31_pU6.opt-gRNA$^{eGFP.9}$	#183	ACCGCGAGGAGCTGTTCACCG
	#184	AAACCGGTGAACAGCTCCTCG
AW46_pU6.opt-gRNA$^{eGFP.13}$	#195	ACCGGGCGAGGAGCTGTTCACCG
	#196	AAACCGGTGAACAGCTCCTCGCC
AX27_pU6.gRNA$^{eGFP.15.c}$	#199	ACCGATGCCGTTCTTCTGCTTGT
	#200	AAACACAAGCAGAAGAACGGCAT
AZ43_pU6.opt-gRNA$^{eGFP.16}$ and AX03_pU6.gRNA$^{eGFP.16}$	#203	ACCGCTCGTGACCACCCTGACCTA
	#204	AAACTAGGTCAGGGTGGTCACGAG
AB69_pU6.opt-gRNA$^{eGFP.23}$	#377	ACCGCAAGCAGAAGAACGGCATCA
	#378	AAACTGATGCCGTTCTTCTGCTTG
AM28_pU6.opt-gRNA$^{eGFP.25}$	#389	ACCGGACCATGGTGAGCAAGGGCG
	#390	AAACCGCCCTTGCTCACCATGGTC
AM31_pU6.opt-gRNA$^{eGFP.26}$	#395	ACCGGCTCACCATGGTGGCGAC
	#396	AAACGTCGCCACCATGGTGAGC
AK65_pU6.opt-gRNA$^{eGFP.29}$	#521	ACCGATGCCCTTCAGCTCGATG
	#522	AAACCATCGAGCTGAAGGGCAT
AK66_pU6.opt-gRNA$^{eGFP.30}$	#523	ACCGCCGTCGTCCTTGAAGAAGA
	#524	AAACTCTTCTTCAAGGACGACGG
AF69_pU6.opt-gRNA$^{eGFP.31}$	#429	ACCGCACCATGGTGAGCAAGGGCG
	#430	AAACCGCCCTTGCTCACCATGGTG
BF50_pU6.opt-gRNA$^{eGFP.32}$	#691	ACCGGCCGTAGGTCAGGGTGGTCACG
	#692	AAACCGTGACCACCCTGACCTACGGC
BH40_pU6.opt-gRNA$^{eGFP.33}$	#705	ACCGTTGCTCACCATGGTGGCGAC
	#706	AAACGTCGCCACCATGGTGAGCAA
X63_pU6.opt-gRNA$^{CD81.1}$.MCS	#722	ACCGAGAGCGAGCGCGCAACGG
	#723	AAACCCGTTGCGCGCTCGCTCT
X68_pU6.opt-gRNA$^{CD81.2}$.MCS	#724	ACCGGCCTGGCAGGATGCGCGG
	#725	AAACCCGCGCATCCTGCCAGGC
BG30_pU6.opt-gRNA$^{CD81.5}$	#740	ACCGCGCACCCATCACCACCACAG
	#741	AAACCTGTGGTGGTGATGGGTGCG
BG31_pU6.opt-gRNA$^{CD81.6}$	#742	ACCGAAGCAGCAGTCCGGAATCCG
	#743	AAACCGGATTCCGGACTGCTGCTT
BG32_pU6.opt-gRNA$^{CD81.7}$	#744	ACCGCTCATGGGGGCGGGGCGCC
	#745	AAACGGCGCCCCGCCCCCATGAG
BG33_pU6.opt-gRNA$^{CD81.8}$	#746	ACCGCGCAGATTGGAGAGTGAGCT
	#747	AAACAGCTCACTCTCCAATCTGCG
BH53_pU6.opt-gRNA$^{CD81.9}$	#809	ACCGGAGTTGATGCCACAGTGGT
	#810	AAACACCACTGTGGCATCAACTC
BH54_pU6.opt-gRNA$^{CD81.10}$	#811	ACCGATCTGGGAGGGCTCCCCAA
	#812	AAACTTGGGGAGCCCTCCCAGAT
BH55_pU6.opt-gRNA$^{CD81.11}$	#813	ACCGCCAAGGTGCTTGAGGGAGGG
	#814	AAACCCCTCCCTCAAGCACCTTGG
BH56_pU6.opt-gRNA$^{CD81.12}$	#815	ACCGCTGGGAGGGCTCCCCAAAGG
	#816	AAACCCTTTGGGGAGCCCTCCCAG
BH57_pU6.opt-gRNA$^{CD81.13}$	#817	ACCGCAGTACTTATAGGGCGCCG
	#818	AAACCGGCGCCCTATAAGTACTG

Supplementary Table S2. Oligonucleotides used for pegRNA assembly.

Plasmid names	Codes	Oligonucleotide sequences (5' → 3')
S77_pU6.PEgRNA^{TLR}	#623	CACCGTAACAGGGTAATGTCGAGGCGTTTT
	#624	CTCTAAAACGCCTCGACATTACCCTGTTAC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#1700	GTGCTTCAGCGTGTCCGGCCATCGACATTACCCTG
	#1701	AAAACAGGGTAATGTCGATGGCCGGACACGCTGAA
AG06_pU6.PEgRNA.2	#701	CACCGTCGCCCTCGAACTTCACCTGTTTT
	#702	CTCTAAAACAGGTGAAGTTCGAGGGCGAC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#1880	GTGCACCCGCGCCGAGGAATGAAGTTCGAGG
	#1881	AAAACCTCGAACTTCATTCCTCGGCGCGGGT
AG07_pU6.PEgRNA.34	#703	CACCGCTCGCCCTTGCTCACCATGGGTTTT
	#704	CTCTAAAACCCATGGTGAGCAAGGGCGAGC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#1882	GTGCACCGGTCGCCACCGTGGTGAGCAAG
	#1883	AAAACTTGCTCACCACGGTGGCGACCGGT
S70_pU6.PEgRNA.16	#617	CACCGCTCGTGACCACCCTGACCTAGTTTT
	#618	CTCTAAAACTAGGTCAGGGTGGTCACGAGC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#1418	GTGCAAGCACTGCACGCCGTGGGTCAGGGTGGTCA
	#1419	AAAATGACCACCCTGACCCACGGCGTGCAGTGCTT
BG26_pU6.PEgRNA$^{CD81.3}$	#748	CACCGCCCGGCCGCCCCTCAGCTAGTTTT
	#749	CTCTAAAACTAGCTGAGGGGCGGCCGGGC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#1949	GTGCCCCTCGTTAGCTGAGGGGCGG
	#1950	AAAACCGCCCCTCAGCTAACGAGGG
BG27_pU6.PEgRNA$^{CD81.4}$	#750	CACCGCATCAAGAGCCGCCGCCCCGTTTT
	#751	CTCTAAAACGGGGCGGCGGCTCTTGATGC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#1951	GTGCCCATCGAGGGGCGGCGG
	#1952	AAAACCGCCGCCCCTCGATGG
BH58_pU6.PEgRNA$^{CD81.9}$	#819	CACCGGAGTTGATGCCACAGTGGTGTTTT
	#820	CTCTAAAACACCACTGTGGCATCAACTCC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTA ACTTGCTATTTCTAG
	#2076	GTGCGCAAAGCGAACCACTGTGGCA
	#2077	AAAATGCCACAGTGGTTCGCTTTGC
BH59_pU6.PEgRNA$^{CD81.10}$	#821	CACCGATCTGGGAGGGCTCCCCAAGTTTT
	#822	CTCTAAAACTTGGGGAGCCCTCCCAGATC

	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG
	#2078	GTGCACCTCGTTTGGGGAGCCCT
	#2079	AAAAAGGGCTCCCCAAACGAGGT
BH61_pU6.PEgRNA[CD81.12]	#825	CACCGCTGGGAGGGCTCCCCAAAGGGTTTT
	#826	CTCTAAAACCCTTTGGGGAGCCCTCCCAGC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG
	#2082	GTGCCTCACGTCCTTTGGGGAGC
	#2083	AAAAGCTCCCCAAAGGACGTGAG
BH62_pU6.PEgRNA[CD81.13]	#827	CACCGCAGTACTTATAGGGCGCCGGTTTT
	#828	CTCTAAAACCGGCGCCCTATAAGTACTGC
	#1424	AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
	#1425	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG
	#2084	GTGCGGGACGGCGGCGCCCTATA
	#2085	AAAATATAGGGCGCCGCCGTCCC

Note: Green, grey and magenta oligonucleotides (sense and antisense) encode, respectively, sequence-specific spacers, pegRNA scaffolds and pegRNA 3' extensions with PBS and RT sequences.

Supplementary Table S3. Transfection scheme for assessing the activities of prime editors in HER.TLR[TetO.KRAB] cells (**Figure 2B, left panel**).

HER.TLR[TetO.KRAB] cells	4.0 ×10[5] cells per well of 24-well plates						
	2000 ng DNA and 9.60 µl PEI (1 mg ml[-1]) per well (medium replaced at 6 h post-transfection)						
Encoded products	Cas9[H840A]	PE2	pegRNA.TLR	gRNA.8	gRNA.16	gI-SceI	EGFP
Plasmid codes	AT79	S65	S77	AW24	AX03	AM51	C55
Construct length (bp)	9215	113	2305	3047	3047	3056	5771
1	1086		359				555
2		108	359				555
3	1086		359	119			555
4		108	359	119			555
5	1086		359		119		555
6		108	359		119		555
7	1086					359	555
8		108				359	555

Supplementary Table S4. Transfection scheme for testing the activities of prime editors in HER.TLR[KRAB] cells (**Figure 2B, right panel**).

HER.TLR[KRAB] cells	4.5 ×10[5] cells per well of 24-well plates						
	2000 ng DNA and 9.60 µl PEI (1 mg ml[-1]) per well (medium replaced at 6 h post-transfection)						
Encoded products	Cas9[H840A]	PE2	pegRNA.TLR	gRNA.8	gRNA.16	gI-SceI	EGFP
Plasmid codes	AT79	S65	S77	AW24	AX03	AM51	C55
Construct length	9215	11389	2305	3047	3047	3056	5771
1	1086		359				555
2		1086	359				555
3	1086		359	119			555
4		1086	359	119			555
5	1086		359		119		555
6		1086	359		119		555
7	1086					359	555
8		1086				359	555

Supplementary Table S5. Transfection scheme for determining target gene knockout frequencies induced by prime editors in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 2D**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.5 ×10^5 cells per well of 24-well plates									
	1100 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)									
Encoded products	Cas9 ^{H840A}	PE2	pegRNA.2	gRNA. 21	gRNA. T2	pegRNA.34	gRNA. 6	gRNA. 8	gI-SceI	DsRed
Plasmid codes	AT79	S65	AG06	BB11	AT44	AG07	AW20	AW24	AM51	AM37
Construct	9215	11389	2306	3047	3974	2305	3046	3047	3056	4712
1	750								250	100
2		750	250							100
3		750	250	100						100
4		750	250		100					100
5		750				250				100
6		750				250	100			100
7		750				250		100		100
8		750							250	100

Supplementary Table S6. Transfection scheme for determining target gene knockout frequencies induced by prime editors in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 3B**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.5 ×10^5 cells per well of 24-well plates						
	1100 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)						
Encoded products	Cas9^{H840A}	PE2	pegRNA.16	gRNA.2	gRNA.7	gI-SceI	DsRed
Plasmid codes	AT79	S65	S70	AV59	AW22	AM51	AM37
Construct length (bp)	9215	11389	2311	3047	3047	3056	4712
1		750	250			250	100
2	750		250	100			100
3		750	250	100			100
4	750		250		100		100
5		750			100		100
6	750					250	100
7		750				250	100

Supplementary Table S7. Transfection scheme for assessing the impact of chromatin conformations on the performance of gRNAs and pegRNAs in HER.TLR$^{TetO.KRAB}$ cells (**Figure 4A**).

HER.TLR$^{TetO.KRAB}$ cells	4.0 ×10^5 cells per well of 24-well plates			
	2000 ng DNA and 9.60 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)			
Encoded products	Cas9	pegRNA.TLR	gRNA.TLR	EGFP
Plasmid codes	AV62	S77	AW24	C55
Construct length (bp)	9215	2305	3047	5771
1	1086	359		555
2	1086		359	555

Supplementary Table S8. Transfection scheme for testing the impact of chromatin conformations on the performance of gRNAs and pegRNAs in HER.TLR^{KRAB} cells (**Figure 4B**).

HER.TLR^{KRAB} cells	4.5 ×10^5 cells per well of 24-well plates			
	2000 ng DNA and 9.60 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)			
Encoded products	Cas9	pegRNA.TLR	gRNA.TLR	EGFP
Plasmid codes	AV62	S77	AW24	C55
Construct length (bp)	9215	2305	3047	5771
1	1086	359		555
2	1086		359	555

Supplementary Table S9. Transfection scheme for assessing the impact of chromatin conformations on the performance of gRNAs and pegRNAs in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 4C**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.5 ×10^5 cells per well of 24-well plates					
	1000 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)					
Encoded products	Cas9	pegRNA.16	gRNA.16	pegRNA.2	gRNA.2	DsRed
Plasmid codes	AV62	S70	AX03	AG06	AV59	AM37
Construct length (bp)	9215	2311	3047	2306	3047	4712

1	750	250				100
2	750		250			100
3	750			250		100
4	750				250	100

Supplementary Table S10. Transfection scheme for inducing closed chromatin remodeling at *CD81* loci in HEK293T cells. (**Supplementary Figure S4**).

HEK293T cells	2.0 ×10^5 cells per well of 24-well plates			
	1200 ng DNA and 5.27 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)			
Encoded products	CRISPRoff v2.1	gRNA.CD81.1	gRNA.CD81.2	DsRed
Plasmid codes	W57	X63	X68	AM37
Construct length (bp)	11885	2311	3047	4712
1	947.3	252.7		
2	947.3		252.7	
3	782.5	208.8	208.8	
4	947.3			252.7
5				1200

Supplementary Table S11. Transfection scheme for assessing the activities of prime editors at *CD81* loci in open and closed chromatin in HEK293T cells (**Figure 5F**).

HEK293T cells (CD81-negative or CD81-positive)	2.5 ×10^5 cells per well of 24-well plates						
	1100 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)						
Encoded products	PE2	pegCD81.3	pegCD81.4	gCD81.5	gCD81.6	gCD81.7	gCD81.8
Plasmid codes	S65	BG26	BG27	BG30	BG31	BG32	BG33
Construct length (bp)	11389	2300	2296	3057	3057	3056	3057
1	750	250					
2	750	250		100			
3	750	250			100		
4	750	250				100	
5	750		250				
6	750		250				100

HEK293T cells (CD81-negative or CD81-positive)	2.0 ×10^5 cells per well of 24-well plates				
	1100 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)				
Encoded products	PE2	pegCD81.9	pegCD81.10	gCD81.12	gCD81.13
Plasmid codes	S65	BH53	BH54	BH56	3172
Construct length (bp)	11389	3172	3172	3173	3057
1	915	250			
2	750	250		100	
3	750	250			100
4	750	250			
5	750		250		
6	750		250		

Supplementary Table S12. Transfection scheme for determining editing frequencies induced by the adenine base editor ABEmax in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 6A**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.0 ×10^5 cells per well of 24-well plates							
	750 ng DNA and 3.29 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)							
Encoded products	Cas9	ABEmax	gRNA.2	gRNA.16	gRNA.31	gRNA.32	gI-SceI	DsRed
Plasmid codes	AV62	BD09	BF23	AZ43	AF69	BF50	AM51	AM37
Construct length (bp)	9215	10522	3172	3048	3046	3047	3056	4712
1	418.0			138.2				193.8
2		431.0					125.2	193.8
3		427.4	128.8					193.8
4		431.3		124.9				193.8
5		431.3			124.9			193.8
6		431.3				124.9		193.8

Supplementary Table S13. Transfection scheme for determining editing frequencies induced by the cytidine base editor coBE3-2NLS in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 6A**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.0 ×10^5 cells per well of 24-well plates								
	750 ng DNA and 3.29 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)								
Encoded products	Cas9	coBE3-2NLS	gRNA.2	gRNA. 16	gRNA. 26	gRNA. 29	gRNA. 30	gI-SceI	DsRed
Plasmid codes	AV62	BC58	BF23	AZ43	AM31	AK65	AK66	AM51	AM37
Construct	9215	10870	3172	3048	3055	3055	3056	3056	4712
1	417.6			138.6					193.8
2		434.1						122.1	193.8
3		430.6	125.6						193.8
4		434.5							193.8
5		434.1		122.1					193.8
6		434.2			122				193.8
7		434.2				122			193.8
8		434.1					122.1		193.8

Supplementary Table S14. Transfection scheme for determining editing frequencies induced by the adenine base editor ABEmax in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 7**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.0 ×10^5 cells per well of 24-well plates								
	1100 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)								
Encoded products	ABEmax	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	DsRed
Plasmid codes	BD09	BA21	AW18	AW20	AW22	AW31	AW46	AZ43	AM37
Construct length (bp)	10522	3056	3045	3046	3046	3044	3046	3057	4712
1	682.0	198.0							220.0
2	682.0		198.0						220.0
3	682.0			198.0					220.0
4	682.0				198.0				220.0
5	682.0					198.0			220.0
6	682.0						198.0		220.0
7	682.0							198.0	220.0

Encoded products	ABEmax	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gI-SceI	DsRed
Plasmid codes	BD09	AB69	AM28	AM31	AF69	BF50	BH40	AM51	AM37
Construct length (bp)	10522	3057	3057	3055	3057	3157	3173	3056	4712
8	682.0	198.0							220.0
9	682.0		198.0						220.0
10	682.0			198.0					220.0
11	682.0				198.0				220.0
12	682.0					198.0			220.0
13	682.0						198.0		220.0
14	682.0							198.0	220.0

Supplementary Table S15. Transfection scheme for determining editing frequencies induced by the cytidine base editor coBE3-2NLS in HEK.EGFP$^{TetO.KRAB}$ cells (**Figure 7**).

HEK.EGFP$^{TetO.KRAB}$ cells	2.0 ×10^5 cells per well of 24-well plates									
	1100 ng DNA and 4.61 µl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)									
Encoded products	coBE3-2NLS	gRNA. 2	gRNA.3	gRNA. 5	gRNA. 6	gRNA. 7	gRNA. 9	gRNA. 13	gRNA. 15C	DsRed
Plasmid codes	BC58	BF23	BA21	AW18	AW20	AW22	AW31	AW46	AX27	AM37
Construct length (bp)	10870	3172	3056	3045	3046	3046	3044	3046	3046	4712
1	687.0	193.0								220.0
2	687.0		193.0							220.0
3	687.0			193.0						220.0
4	687.0				193.0					220.0
5	687.0					193.0				220.0
6	687.0						193.0			220.0

7	687.0							193.0		220.0
8	687.0								193.0	220.0
Encoded	coBE3-	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gRNA.	gI-SceI	DsRed
Plasmid codes	BC58	AZ43	AB69	AM28	AM31	AF69	BF50	BH40	AM51	AM37
Construct length	10870	3057	3057	3057	3055	3057	3157	3173	3056	4712
9	687.0	193.0								220.0
10	687.0		193.0							220.0
11	687.0			193.0						220.0
12	687.0				193.0					220.0
13	687.0					193.0				220.0
14	687.0						193.0			220.0
15	687.0							193.0		220.0
16	687.0								193.0	220.0

Supplementary Table S16. Transfection scheme for assessing the activities of both base editors at *CD81* loci in open and closed chromatin in HEK293T cells (**Figure 8** and **Supplementary Figure S6**).

HEK293T cells (CD81-negative or CD81-positive)	2.0 ×10^5 cells per well of 24-well plates 1000 ng DNA and 4.39 μl PEI (1 mg ml^{-1}) per well (medium replaced at 6 h post-transfection)							
Encoded products	ABEmax	coBE3-2NLS	gCD81.9	gCD81.10	gCD81.11	gCD81.12	gCD81.13	gI-SceI
Plasmid codes	BD09	BC58	BH53	BH54	BH55	BH56	BH57	AM51
Construct length	10522	10870	3172	3173	3173	3173	3172	3056
1	845.0		255.0					
2	845.0				255.0			
3	845.0					255.0		
4	845.0						255.0	
5	845.0							255.0
6		852.0	248.0					
7		852.0		248.0				
8		852.0			248.0			
9		852.0				248.0		
10		852.0					248.0	
11		852.0						248.0

Supplementary Table S17. Composition of mixtures used for qPCR amplification.

Targets	**Primer codes**	**Primers (5' → 3')**	**SYBR Green Master mix**	**Primers (μM)**	**Amplicons size (bp)**
CD81 (qPCR)	#1958	CTGCTTTGACCACCTCAGTGCT	1×	0.2	798
	#1959	TGGCAGCAATGCCGATGAGGTA			
GAPDH (qPCR)	#119	AGCCACATCGCTCAGACACC	1×	0.2	302
	#120	GTACTCAGCGCCAGCATCG			
CD81 b (ChIP-qPCR)	#2004	ATCAACTCCTTCAGGAAGCCC	1×	0.2	113
	#2005	CCGGGAGAACAACCCATTCC			
CD81 c (ChIP-qPCR)	#2006	CAGCAATTCTCCCCTTCCGT	1×	0.2	120
	#2007	TTGCTCACATTGCTCTCCGG			
GAPDH a (ChIP-qPCR)	#1998	CGCGCCCCCGGTTTCTAT	1×	0.2	80
	#1999	GATGCGGCTGACTGTCGAA			
GAPDH b (ChIP-qPCR)	#2000	TACTAGCGGTTTTACGGGCG	1×	0.2	166
	#2001	TCGAACAGGAGGAGCAGAGAGCGA			
GAPDH c (ChIP-qPCR)	#2024	TAGGCGCTCACTGTTCTCTC	1×	0.2	82
	#2025	CGTTGACTCCGACCTTCAC			
ACTB (ChIP-qPCR)	#2020	AACTCTCCCTCCTCCTCTTCC	1×	0.2	69
	#2021	CCTCTCCCCTCCTTTTGC			
ZNF184 (ChIP-qPCR)	#2022	TTGGGAATATGAAGGCAGTT	1×	0.2	60
	#2023	TCCTTTGGCAGTGTCTGTTG			

Supplementary Table S18. Thermocycler program used in qPCR amplification.

Steps	**Temperatures**	**Times**
Initial denaturation	95.0 ℃	5 min
Denaturation	95.0 ℃	10 sec

Annealing	60.0 ℃	30 sec
Elongation		
Plate read		
Cycles (Go to step 2)	45	
Melt curve analysis	65.0 ℃ to 95.0 ℃ (increase in 0.5 ℃ increments with a hold time of 5 sec	
Plate read		

Supplementary Table S19. Gene-specific primer sequences and concentrations used in the NGS analyses.

Target	Primer code	Primers (5' → 3') / final concentrations (μM)
eGFP	#1791	GATGTGTATAAGAGACAGGCACGACTTCTTCAAGTCCG / 0.5
	#1792	CGTGTGCTCTTCCGATCTAGTTCACCTTGATGCCGTTC / 0.5
eGFP	#1884	GATGTGTATAAGAGACAGATGCCACCTACGGCAAGCTG / 0.5
	#1885	CGTGTGCTCTTCCGATCTCCTCCTTGAAGTCGATGCCC / 0.5
eGFP	#1916	GATGTGTATAAGAGACAGCGATCACGAGACTAGCCTCG / 0.5
	#1917	CGTGTGCTCTTCCGATCTTAGGTCAGGGTGGTCACGAG / 0.5
eGFP	#2087	GATGTGTATAAGAGACAGCATCGACTTCAAGGAGGACGG / 0.5
	#2088	CGTGTGCTCTTCCGATCTGGGTGTTCTGCTGGTAGTGG / 0.5
CD81	#1987	GATGTGTATAAGAGACAGTTTCGGGGCCTCTGTGCTCG / 0.5
	#1988	CGTGTGCTCTTCCGATCTACCTCCGGCAAAGTGTGCGC / 0.5
CD81	#1989	GATGTGTATAAGAGACAGGGATTCCGGACTGCTGCTTCGC / 0.5
	#1990	CGTGTGCTCTTCCGATCTACCCCAGCTTCTGGGCCATC / 0.5
CD81	#2095	GATGTGTATAAGAGACAGGGTGCAGCGACCCCATACCCC / 0.5
	#2096	CGTGTGCTCTTCCGATCTGCCTGGCAGGATGCGCGGTG / 0.5

Supplementary Table S20. PCR cycling parameters used in the NGS analyses.

Target	Initial denaturation	Denaturation	Annealing	Elongation	Cycles	Final elongation
eGFP (1791+1792) (gene-specific PCR)	98 ℃	98 ℃	61.1 ℃	72 ℃	35	72 ℃
	30 sec	10 sec	10 sec	10 sec		5 min
eGFP (1884+1885) (gene-specific PCR)	98 ℃	98 ℃	70.5 ℃	72 ℃	35	72 ℃
	30 sec	10 sec	10 sec	10 sec		5 min
eGFP (1916+1917) (gene-specific PCR)	98 ℃	98 ℃	67.6 ℃	72 ℃	35	72 ℃
	30 sec	10 sec	10 sec	10 sec		5 min
eGFP (2087+2088) (gene-specific PCR)	98 ℃	98 ℃	67.0 ℃	72 ℃	35	72 ℃
	5 min	10 sec	10 sec	10 sec		5 min
CD81 (1987+1988) (gene-specific PCR)	98 ℃	98 ℃	67.6 ℃	72 ℃	35	72 ℃
	30 sec	10 sec	10 sec	10 sec		5 min
CD81 (1989+1990) (gene-specific PCR)	98 ℃	98 ℃	71.6 ℃	72 ℃	35	72 ℃
	30 sec	10 sec	10 sec	10 sec		3 min
CD81 (2095+2096) (gene-specific PCR)	98 ℃	98 ℃	71.0 ℃	72 ℃	35	72 ℃
	5 min	10 sec	10 sec	10 sec		5 min
Barcode PCR	98 ℃	98 ℃	62.0 ℃	72 ℃	10	72 ℃
	30 sec	10 sec	10 sec	10 sec		3 min

Supplementary Table S21. Composition of PCR mixtures used in gene-specific amplifications for NGS analyses.

Component	Volume	Final Concentration
5× Phusion HF Buffer	4 μl	1×
dNTPs (2.5 mM each)	1.6 μl	0.2 mM (each)
PCR Grade Water	10.7 μl	-
Forward primer (10 μM)	1 μl	0.5 μM
Reverse primer (10 μM)	1 μl	0.5 μM
gDNA	1.5 μl	-
Phusion DNA Polymerase (2 U/μl)	0.2 μl	0.02 U/μl
Total reaction volume	20 μl	-

Supplementary Table S22. Barcoded PCR primers used in the NGS analyses.

Primer	Primers (5' → 3') / final concentrations (0.25 µM)
Fun-i501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Fun-i503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Fun-i504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i517	AATGATACGGCGACCACCGAGATCTACACGCGTAAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i501D	AATGATACGGCGACCACCGAGATCTACACTATAGCCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i502D	AATGATACGGCGACCACCGAGATCTACACATAGAGGCTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i503D	AATGATACGGCGACCACCGAGATCTACACCCTATCCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i504D	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i505D	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i506D	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i507D	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAGATGTGTATAAGAGACA
Fun-i701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i709	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Fun-i712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Supplementary Table S23. Composition of PCR mixtures for barcoded PCR amplification in the NGS analyses.

Component	**Volume**	**Final Concentration**
5× Phusion HF Buffer	4 µl	1×
dNTPs (2.5 mM each)	1.2 µl	0.15 mM (each)
PCR Grade Water	11.6 µl	-
Index primer p5-XX (5 µM)	1 µl	0.25 µM
Index primer p7-XX (5 µM)	1 µl	0.25 µM
Purified PCR product	1 µl	-
Phusion DNA Polymerase (2 U/µl)	0.2 µl	0.02 U/µl
Total reaction volume	20 µl	-

Genome editing technologies permit introducing specific genetic changes within the vast genomes of living eukaryotic cells *in vitro* and *in vivo*. As such, these technologies are having an ever-increasing impact on both basic and applied science. In the framework of human health, the ultimate goal is that of translating these techniques into therapeutically relevant applications, including those directed at (i) permanently correcting *ex vivo* or *in vivo* mutations associated with hereditary diseases, (ii) interfering with the replication cycle of infectious agents, and (iii) improving the efficacy and safety of cancer immunotherapies that make use of genetically engineered T lymphocytes or natural killer cells as armed "living drugs". Notwithstanding the remarkable progress observed during the past two decades on the development of genome editing tools and strategies, trend fostered after the inception in 2013 of RNA-guide nucleases (RGNs), several longstanding bottlenecks limit the application of these technologies as effective and safe gene and cell therapies. These bottlenecks include large-scale and small-scale mutagenic events (stochastic or otherwise), off-target activities, activation of DNA damage responses and ineffective cellular delivery of the large and multiple reagents required to effectuate the intended chromosomal modification(s) in the proper cell types, tissue or organs.

Typically, genome editing protocols comprise the delivery of sequence-tailored designer nucleases (e.g., CRISPR-Cas9-based RGNs) that, upon targeted double-stranded DNA break (DSB) formation and ensuing activation of endogenous DNA repair pathways, yield specific chromosomal DNA modifications. For the purpose of site-specific chromosomal addition of exogenous genetic information (gene targeting or knock-in), delivery of designer nucleases is combined with that of surrogate donor DNA-repairing templates whose sharing of homology to genomic target sequences, makes them prone to precise homology-directed DNA repair (HDR) processes. Yet, instead of HDR-mediated genome editing, designer nuclease-induced DSBs are more often engaged by competing error-prone DNA repair mechanisms, e.g., non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). Although numerous studies have shown that HDR-mediated genome editing can be favored through the inhibition of NHEJ and MMEJ factors, it is, in principle, preferable to improve the ectopic HDR process itself rather than interfering with the activity of DNA repair factors whose consequences for genomic stability are mostly unpredictable. Another critical bottleneck concerns the need for effective and safe delivery of the large and multicomponent elements underpinning genome editing procedures. Hence, **Chapter 1**, besides reviewing classical and more recent genome editing tools and strategies, it also covers the use of adenoviral vectors (AdVs) as delivery agents for targeted genetic manipulation of human stem cells, progenitor cells, and their differentiated progenies, focusing on *in vitro* and *ex vivo* protocols. In this context, high-capacity adenoviral vectors (HC-AdVs) deleted of all viral genes constitute particularly valuable vehicles for ferrying large genome editing reagents owing to their low cytotoxicity profile and amenability to cell tropism modifications. Indeed, in this thesis, HC-AdVs displaying CD46-specific capsid fibers from adenovirus serotype-50 instead of coxsackievirus and adenovirus receptor (CAR)-binding fibers from prototypic serotype-5, permitted effective testing of emerging genome editing principles in scientifically and therapeutically relevant CAR-negative human cell types, e.g., mesenchymal stem cells (hMSCs) and muscle progenitor cells (myoblasts) as well as CD46- and CAR-positive induced pluripotent stem cell (iPSC)-derived cardiomyocytes.

Motivated by the aforementioned limitations of commonly used genome editing procedures, strongly associated with their dependency on designer nucleases, the experimental chapters presented in this thesis focus on investigating genome editing principles based on the use of sequence- and strand-specific nucleases ("nickases"). In this regard, **Chapter 2** demonstrates that *in trans* paired nicking (ITPN), comprising simultaneous single-stranded DNA break (SSB) formation at genomic target sites and donor DNA constructs by *Streptococcus pyogenes* CRISPR-Cas9-derived nickases, triggers seamless and scarless HDR-mediated gene knock-ins at endogenous loci whose products are essential for regular cell function, in particular, alleles fundamental for DNA damage responses (i.e., *H2AX* and *PARP1*) and for the maintenance of pluripotency in *bona fide* pluripotent stem cells and iPSCs (i.e., *OCT4*). Importantly, the low mutagenic character of ITPN was shown to preserve target protein dosages and to prevent phenotypic and fitness losses in gene-edited cell populations. In addition, through a collaborative effort, **Chapter 2** introduces the orthogonal high-throughput genome-wide translocation sequencing (oHTGTS) technique for unbiased identification and characterization of off-target sites and effects, respectively, resulting from cleaving versus nicking RGNs. Using oHTGTS it was established that nicking RGNs greatly reduce the frequency of large-scale chromosomal rearrangements and translocations when compared with their DNA cleaving counterparts. Nicking RGNs could nonetheless

induce detectable translocations involving on-target and off-target sites. Presumably such events can arise when, for example, an advancing replication fork collapses after hitting a nickase-induced SSB product. Moreover, given the fact that eukaryotic genomes contain multiple repetitive elements whose individual units share full or high sequence identity with units scattered elsewhere throughout the genome (e.g., retroelements, amplified gene clusters, gene paralogs and pseudogenes), there is a pressing need to identify high-specificity nickases permitting a judicious access to specific chromosomal sequences while averting similar off-target sites. Therefore, in **Chapter 3**, a representative panel of RuvC-disabled *S. pyogenes* Cas9 nickases (SpCas9^{D10A}) was assembled on the basis of the respective high-specificity nucleases, i.e., SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A}, Sniper-Cas9^{D10A}, xCas9-3.7^{D10A}, evoCas9^{D10A} and SpCas9-HF1^{D10A}. Subsequent benchmarking experiments and functional screens described in **Chapter 3** identify high-specificity SpCas9^{D10A} variants that can outperform their regular counterparts at the levels of discriminating on-target from off-target sequences and minimizing genome-wide translocations as determined through functional screens and oHTGTS analysis, respectively. Moreover, high-specificity SpCas9^{D10A} nickases operating as dual nicking RGNs also outperformed their conventional counterparts in terms of yielding highly specific gene knockouts and, together with matched donor constructs, achieve specific gene knock-ins by minimizing off-target insertions at similar pseudogene elements. Following from these findings, **Chapter 4** illustrates that high-specificity SpCas9^{D10A} nickases are capable of eliciting ITPN genome editing to the same or higher extents than those triggered by the parental SpCas9^{D10A} protein, including at “safe harbor” loci (e.g., *AAVS1* and *CCR5*) whose HDR-mediated DNA targeting allows for long-term and homogenous transgene expression in engineered cell populations. Critically, **Chapter 4** further shows that, in contrast to regular and high-specificity SpCas9 nucleases, neither regular nor high-specificity SpCas9^{D10A} nickases activate the canonical P53-dependent DNA damage response signaling pathway in human iPSCs, further stressing the potentially higher safety profile of nickases over nucleases for the genomic engineering of cells with high sensitivity to DNA damage, e.g., pluripotent and tissue-specific stem cells. Indeed, these data indicate that SpCas9^{D10A} nickases might offer a heightened safety profile to engineered cell products derived from stem cells as, in addition to cell-cycle arrest and apoptosis, DSB-induced signaling pathways have been associated with the selection of cells bearing mutations in cancer-associated genes, e.g., *TP53* itself and *KRAS*.

As aforementioned, a critical bottleneck regarding the application of genome editing technologies concerns the need for introducing, in an effective and non-cytotoxic manner, the required large and multicomponent reagents into cells, tissues or organs of interest. **Chapter 5** demonstrates that HC-AdVs, in particular CD46-targeting HC-AdVs, are a suitable option for all-in-one delivery of full-length prime editing reagents, in the form of prime editors and prime editing gRNAs (pegRNAs), into human cells regardless of their transformation and replication statuses. Indeed, up to 90% prime editing efficiencies are achievable without overt cytotoxicity in transduced cells. Additionally, a direct correlation between the replication status of target cells and prime editing activities was found by using this cell cycle-independent viral vector delivery platform.

The findings presented in **Chapter 5** are further expanded in **Chapter 6** by leveraging HC-AdVs for delivering advanced prime editing systems designed for installing precise *DMD* gene edits in human myogenic cells. In particular, in myoblasts and mesenchymal stem cells, with efficiencies of up to 80% and 64%, respectively, and in cardiomyocytes differentiated from iPSCs isolated from a Duchenne muscular dystrophy (DMD) patient, with efficiencies of up to 82%. Defective *DMD* alleles underlie DMD (OMIM #310200), a common and lethal X-linked muscle-wasting disorder that afflicts circa 1 in 4,700 boys whose treatment options are, currently, merely palliative. HC-AdV transduction experiments designed for defective *DMD* reading frame repair readily led to the detection of mRNA transcripts encoding proteins corresponding to shortened, yet partially functional, dystrophin variants (i.e., Becker-like dystrophins) in unselected muscle cell populations. Crucially, proximity ligation assays revealed that the resulting Becker-like dystrophin proteins were capable of connecting to β-dystroglycan, a key component of the dystrophin-associated glycoprotein complex located at the sarcolemma of normal muscle cells. Moreover, additional *DMD* reading frame restoration experiments demonstrate the feasibility of leveraging HC-AdV delivery for multiplexing prime editing based on the concerted action of pairs of prime editing complexes. Finally, the straightforward HC-AdV delivery process combined with the non-mutagenic character of prime editing can be exploited for the selective accumulation of precise chromosomal edits in target cell populations through consecutive transduction rounds. Taken together,

the research covered in **Chapter 5** and **Chapter 6** reveals that the integrated delivery of prime editing systems in single HC-AdV particles yields efficient and precise modification of target alleles in human stem/progenitor cells. As a consequence, HC-AdV-assisted prime editing warrants further research and testing, including for the modelling and repairing of genetic defects in *ex vivo* and *in vivo* settings.

The chromatin environment of genomic DNA sequences varies in different cells types and, often, is highly dynamic as a result of the spatiotemporal regulation of epigenetic mechanisms underlying organismal development and cellular differentiation. Interestingly, it is becoming evident that the performance of genome-editing reagents is dependent on a combination of genetic and epigenetic variables, i.e., the target nucleotide sequences *per se* and their epigenetically-regulated chromatin environment, respectively. Possibly, besides controlling to varying degrees the accessibility of target sequences to genome editing tools, the local epigenetic context may also influence DNA repair pathway choices and other DNA editing determinants whose combined effects ultimately define the observed genome modification endpoints. For instance, previous studies obtained in the hosting group have revealed that the activity of designer nucleases based on CRISPR-Cas9 systems and transcription activator-like effectors are significantly hampered by heterochromatic states and that the ratio between HDR and mutagenic NHEJ events can vary in a chromatin context-dependent manner.

Considering that base editors and prime editors are formed by fusing CRISPR nickases to secondary effector domains (i.e., deaminases and reverse transcriptases, respectively), in **Chapter 7**, it is investigated whether there are cause-effect associations between alternate chromatin conformations and genome editing endpoints when using these two powerful DSB-independent genome editing platforms. By implementing complementary loss-of-function and gain-of-function cellular systems, prime editing was found to be frequently hindered at heterochromatin impinged by the KRAB/KAP-1/HP1 axis alone or together with the DNA methyltransferases DNMT3A and DNMT3L. Moreover, the extended portions of gRNAs forming pegRNAs contribute to the underperformance of prime editors at heterochromatic sequences. Notably, in striking contrast with prime editors and designer nucleases (CRISPR-based or otherwise), the DNA editing activity of base editors at closed heterochromatic states ranges in a target site-dependent manner from lower to, often, significantly higher than that observed at open euchromatin. In addition, the chromatin environment of target sequences was also found to be capable of significantly influencing the fidelity and purity of base editing products in a gRNA-dependent manner. As a corollary, the performance and safety profiles of base editing and prime editing technologies necessitates in-depth scrutiny for guiding their selection, further development and application in specific contexts. On the basis of the data presented in **Chapter 7**, one can also submit that algorithms trained to predict the activities of base editing and prime editing reagents, besides inputs on target sequences, will profit from the processing of information regarding the epigenetic context of said sequences. Finally, these data can further guide the development of combinatorial strategies in which targeted epigenetic modulators and DSB-free genome editing tools act in concert for achieving a more efficient and/or more precise genetic modification of cellular (epi)genomes.

In summary, by predominantly investigating genome editing tools and strategies based on CRISPR-Cas9 nickases as such or on their prime editing and base editing derivatives, this thesis provides insights on how these tools and strategies operate in human cells opening up, in this process, new avenues for the seamless modification of cellular (epi)genomes. Moving ahead it is expected that by further developing and refining “soft” genome editing procedures that, besides the efficiency, take into account specificity and accuracy parameters, will allow for translating ‘genomic surgery’ interventions into effective and safe gene and cell therapies.

Addendum

Nederlandese Samenvatting / ***214***

List of Publications / ***218***

Curriculum Vitae / ***220***

Acknowledgments / ***221***

Nederlandse Samenvatting

Technologieën voor genoombewerking maken het mogelijk om in *vitro* en *in vivo* specifieke genetische veranderingen aan te brengen in de enorme genomen van levende eukaryotische cellen. Als zodanig hebben deze technologieën een steeds grotere impact op zowel de fundamentele als de toegepaste wetenschap. In het kader van de menselijke gezondheid is het uiteindelijke doel om deze technieken te vertalen naar therapeutisch relevante toepassingen, waaronder (i) het permanent *ex vivo* of *in vivo* corrigeren van mutaties die geassocieerd zijn met erfelijke ziekten, (ii) het verstoren van de replicatie van infectieuze agentia, en (iii) het verbeteren van de werkzaamheid en veiligheid van kankerimmunotherapieën die gebruik maken van genetisch gemodificeerde T-lymfocyten of natuurlijke killercellen als gewapende 'levende geneesmiddelen'. Ondanks de opmerkelijke vooruitgang die de afgelopen twintig jaar is geboekt op het gebied van de ontwikkeling van genoombewerkingstools en -strategieën, een trend die is bevorderd na de introductie in 2013 van RNA-gidsnucleasen (RGNs), zijn er verschillende al lang bestaande knelpunten die de toepassing van deze technologieën in effectieve en veilige gen- en celtherapieën beperken. Deze knelpunten omvatten grote en kleine mutagene gebeurtenissen (stochastisch of anderszins), off-target activiteiten, activering van DNA-schaderespons en ineffectieve cellulaire afgifte van de grote en meerdere reagentia die nodig zijn om de beoogde chromosomale modificatie(s) in de juiste celtypen, weefsels of organen te bewerkstelligen.

Doorgaans omvatten genoombewerkingsprotocollen de aflevering van op doelsequenties afgestemde designernucleasen (bijv. CRISPR-Cas9-gebaseerde RGNs) die, na gerichte dubbelstrengs DNA-breuk (DSB) vorming en daaropvolgende activering van endogene DNA-reparatieroutes, specifieke chromosomale DNA veranderingen opleveren. Met het oog op plaatsspecifieke chromosomale toevoeging van exogene genetische informatie (gentargeting of knock-in), wordt de toelevering van designernucleasen gecombineerd met die van surrogaatdonor-DNA-herstellende sjablonen, waarvan homologie met genomische doelsequenties hen vatbaar maakt voor nauwkeurige homologiegerichte DNA-reparatieprocessen (HDR). Maar in plaats van HDR-gemedieerde genoombewerking worden door designernuclease geïnduceerde DSBs vaker gebruikt door concurrerende, foutgevoelige DNA-reparatiemechanismen, bijvoorbeeld niet-homologe eindverbinding (NHEJ) en microhomologie-gemedieerde eindverbinding (MMEJ). Hoewel talloze onderzoeken hebben aangetoond dat HDR-gemedieerde genoombewerking kan worden bevoordeeld door de remming van NHEJ- en MMEJ-factoren, verdient het de voorkeur om het HDR-proces zelf te verbeteren, in plaats van de activiteit van DNA-reparatiefactoren te verstoren, waarvan de gevolgen voor genomische stabiliteit meestal onvoorspelbaar is. Een ander kritisch punt betreft de behoefte aan effectieve en veilige afgifte van de grote en uit diverse componenten bestaande elementen, die ten grondslag liggen aan de genoombewerkingsprocedures. Daarom behandelt **Hoofdstuk 1**, naast de beoordeling van klassieke en recentere hulpmiddelen en strategieën voor het bewerken van het genoom, ook het gebruik van adenovirale vectoren (AdVs) als afgiftemiddelen voor gerichte genetische modificatie van menselijke stamcellen, voorlopercellen en hun gedifferentieerde nakomelingen, met de nadruk op *in vitro* en *ex vivo* protocollen. In deze context vormen adenovirale vectoren met hoge capaciteit (HC-AdVs), waarbij alle virale genen zijn verwijderd, bijzonder waardevolle vectoren voor het afgeven van grote genoombewerkingscomponenten vanwege hun lage cytotoxiciteit en hun vatbaarheid voor celtropisme-modificaties. In dit proefschrift maakten HC-AdVs, die CD46-specifieke capsidefibers van adenovirus serotype-50 bevatten in plaats van coxsackievirus en adenovirusreceptor (CAR)-bindende fibers van prototypisch serotype-5, het effectief testen van nieuwe genoombewerkingsmethoden mogelijk op therapeutisch relevante CAR-negatieve menselijke celtypen, zoals mesenchymale stamcellen (hMSCs) en spiervoorlopercellen (myoblasten), maar ook CD46- en hartspiercellen afgeleid van CAR-positieve geïnduceerde pluripotente stamcellen (iPSC).

Gemotiveerd door de bovengenoemde beperkingen van veelgebruikte genoombewerkingsmethoden, sterk geassocieerd met hun afhankelijkheid van designer nucleasen, richten de experimentele hoofdstukken die in dit proefschrift worden gepresenteerd zich op het onderzoeken van genoombewerkingsprincipes gebaseerd op het gebruik van sequentie- en strengspecifieke nucleasen ("nickases"). In dit verband laat **Hoofdstuk 2** zien dat '*in trans* paired nicking' (ITPN), bestaande uit gelijktijdige vorming van enkelstrengige DNA-breuken (SSB) op genomische doelwit-locaties en donor-DNA-constructen door *Streptococcus pyogenes* CRISPR-Cas9-afgeleide nickases, naadloze en littekenloze HDR -gemedieerde gen-knock-ins op endogene loci teweegbrengt, waarvan de producten

essentieel zijn voor de reguliere celfunctie, in het bijzonder allelen die fundamenteel zijn voor reacties op DNA-schade (bijv., *H2AX* en *PARP1*) en voor het behoud van pluripotentie in bonafide pluripotente stamcellen en iPSCs (bijv., *OCT4*). Belangrijk is dat werd aangetoond dat het lage mutagene karakter van ITPN de doeleiwitdoseringen behoudt en fenotypische en fitness-verliezen in gen-bewerkte celpopulaties voorkomt. Daarnaast introduceert **Hoofdstuk 2**, door middel van een samenwerkingsverband, de orthogonale high-throughput genoom-brede translocatie sequencing (oHTGTS) techniek voor identificatie en karakterisering van off-target sites en effecten, respectievelijk resulterend uit knippende versus nicking RGNs. Met behulp van oHTGTS werd vastgesteld dat het induceren van enkelstrengsbreuken door RGNs de frequentie van grootschalige chromosomale herschikkingen en translocaties aanzienlijk vermindert in vergelijking met hun DNA-splitsende tegenhangers. Nicking-RGNs zouden niettemin detecteerbare translocaties kunnen veroorzaken waarbij on-target en off-target sites betrokken zijn. Vermoedelijk kunnen dergelijke gebeurtenissen zich voordoen wanneer bijvoorbeeld een voortschrijdende replicatievork instort na het raken van een door nickase geïnduceerd SSB-product. Bovendien bestaat er, gezien het feit dat eukaryotische genomen vele repetitieve elementen bevatten, waarvan de individuele eenheden een volledige of hoge sequentie-identiteit delen met eenheden die elders in het genoom verspreid zijn (bijvoorbeeld retro-elementen, geamplificeerde genclusters, genparalogen en pseudogenen), een dringende behoefte om nickases met hoge specificiteit te identificeren, die toegang tot specifieke chromosomale sequenties mogelijk maken, terwijl vergelijkbare off-target-sites worden vermeden. Daarom werd in Hoofdstuk 3 een representatief panel van RuvC-uitgeschakeld *S. pyogenes* Cas9 nickases (SpCas9^{D10A}) samengesteld, op basis van betreffende nucleasen met hoge specificiteit, namelijk SpCas9-KA^{D10A}, SpCas9-KARA^{D10A}, eSpCas9(1.1)^{D10A}, Sniper-Cas9^{D10A}, xCas9-3.7^{D10A}, evoCas9^{D10A} en SpCas9-HF1^{D10A}. Daaropvolgende benchmarkingexperimenten en functionele screenings beschreven in **Hoofdstuk 3** identificeren SpCas9^{D10A}-varianten met hoge specificiteit die hun reguliere tegenhangers kunnen overtreffen op het niveau van het onderscheid maken tussen on-target en off-target sequenties en het minimaliseren van genoombrede translocaties zoals bepaald door functionele screens en oHTGTS-analyse. Bovendien presteerden SpCas9^{D10A}-nickases met hoge specificiteit die als 'dual nicking-RGNs' werken ook beter dan hun conventionele tegenhangers in termen van het opleveren van zeer specifieke gen-knock-outs en bereiken ze, samen met gematchte donorconstructen, specifieke gen-knock-ins door off-target-inserties op vergelijkbare pseudogene elementen te minimaliseren. Op basis van deze bevindingen illustreert **Hoofdstuk 4** dat SpCas9^{D10A}-nickases met hoge specificiteit in staat zijn om ITPN-genoombewerking in dezelfde of hogere mate uit te lokken dan die welke worden geactiveerd door het ouderlijke SpCas9^{D10A}-eiwit, inclusief op 'veilige haven'-loci (bijv., *AAVS1* en *CCR5*), waarvan de HDR-gemedieerde DNA-targeting langdurige en homogene transgenexpressie in gemanipuleerde celpopulaties mogelijk maakt. Cruciaal is dat **Hoofdstuk 4** verder laat zien dat, in tegenstelling tot reguliere SpCas9-nucleasen met hoge specificiteit, noch reguliere SpCas9^{D10A}-nickases noch reguliere SpCas9^{D10A}-nickases met hoge specificiteit de canonieke P53-afhankelijke signaalroute voor DNA-schaderespons in menselijke iPSCs activeren, wat het potentieel betere veiligheidsprofiel van nickases boven nucleasen voor de genomische manipulatie van cellen met een hoge gevoeligheid voor DNA-schade, bijvoorbeeld pluripotente en weefselspecifieke stamcellen, verder benadrukt. Deze gegevens duiden er inderdaad op dat SpCas9^{D10A}-nickases een beter veiligheidsprofiel zouden kunnen bieden voor gemodificeerde celproducten die zijn afgeleid van stamcellen, omdat, naast het stoppen van de celcyclus en apoptose, door DSB geïnduceerde signaalroutes in verband zijn gebracht met de selectie van cellen die mutaties in kankergeassocieerde genen, bijvoorbeeld *TP53* zelf en *KRAS,* dragen.

Zoals eerder vermeld, betreft een kritisch knelpunt met betrekking tot de toepassing van genoombewerkingstechnologieën de noodzaak om, op een effectieve en niet-cytotoxische manier, de vereiste grote en uit meerdere componenten bestaande reagentia te introduceren in cellen, weefsels of organen die interessant zijn. **Hoofdstuk 5** laat zien dat HC-AdVs, in het bijzonder op CD46 gerichte HC-AdVs, een geschikte optie zijn voor alles-in-één levering van 'prime editing reagentia' van volledige lengte, in de vorm van prime editors en prime editing gRNAs (pegRNAs), in menselijke cellen, ongeacht hun transformatie- en replicatiestatus. Tot 90% van de primaire 'editing-efficiëntie' is inderdaad haalbaar, zonder duidelijke cytotoxiciteit in getransduceerde cellen. Bovendien werd een directe correlatie gevonden tussen de replicatiestatus van doelcellen en primaire editing activiteiten, door gebruik te maken van dit celcyclus-onafhankelijke platform voor het afleveren van virale vectoren.

De bevindingen gepresenteerd in **Hoofdstuk 5** worden verder uitgebreid in **Hoofdstuk 6** door gebruik te maken van HC-AdVs voor het afleveren van geavanceerde prime editing-systemen die zijn ontworpen voor het installeren van nauwkeurige *DMD*-genbewerkingen in menselijke myogene cellen. In het bijzonder in myoblasten en mesenchymale stamcellen, met efficiënties tot respectievelijk 80% en 64%, en in hartspiercellen gedifferentieerd van iPSCs geïsoleerd uit een patiënt met Duchenne spierdystrofie (DMD), met efficiënties tot 82%. Defecte *DMD*-allelen liggen ten grondslag aan DMD (OMIM #310200), een veel voorkomende en dodelijke X-gebonden spierafbraakstoornis die ongeveer 1 op de 4.700 jongens treft, van wie de behandelingsopties momenteel louter palliatief zijn. HC-AdV-transductie-experimenten ontworpen voor herstel van het defecte *DMD*-leesraam leidden al snel tot de detectie van mRNA-transcripten die coderen voor eiwitten die overeenkomen met verkorte, maar gedeeltelijk functionele, dystrofine-varianten (dwz Becker-achtige dystrofines) in niet-geselecteerde spiercelpopulaties. Cruciaal was dat nabijheids-ligatietesten aantoonden dat de resulterende Becker-achtige dystrofine-eiwitten in staat waren zich te verbinden met β-dystroglycan, een sleutelcomponent van het dystrofine-geassocieerde glycoproteïnecomplex dat zich bevindt in het sarcolemma van normale spiercellen. Bovendien demonstreren aanvullende *DMD*-leesframe herstel-experimenten de haalbaarheid van het gebruik van HC-AdV-levering voor multiplexing van prime-editing op basis van de gecoördineerde actie van paren prime-editingcomplexen. Tenslotte kan het eenvoudige HC-AdV-afgifteproces, gecombineerd met het niet-mutagene karakter van prime editing, worden benut voor de selectieve accumulatie van nauwkeurige chromosomale bewerkingen in doelcelpopulaties door middel van opeenvolgende transductieronden. Alles bij elkaar laat het onderzoek dat in **Hoofdstuk 5** en **Hoofdstuk 6** wordt behandeld zien dat de geïntegreerde levering van ‘prime editing’-systemen in afzonderlijke HC-AdV-deeltjes, efficiënte en nauwkeurige modificatie van doelallelen in menselijke stam-/voorlopercellen oplevert. Als gevolg hiervan rechtvaardigt HC-AdV-ondersteunde ‘prime editing’ verder onderzoek, inclusief voor het modelleren en repareren van genetische defecten in *ex vivo* en *in vivo* omgevingen.

De chromatine-omgeving van genomische DNA-sequenties varieert in verschillende celtypen en is vaak zeer dynamisch als gevolg van de spatiotemporele regulatie van epigenetische mechanismen die ten grondslag liggen aan de ontwikkeling van organismes en cellulaire differentiatie. Interessant genoeg wordt het duidelijk dat de prestaties van reagentia voor het bewerken van het genoom afhankelijk zijn van een combinatie van genetische en epigenetische variabelen, dat wil zeggen respectievelijk de doelnucleotidesequenties als zodanig en hun epigenetisch gereguleerde chromatine-omgeving. Mogelijk kan de lokale epigenetische context, naast het in verschillende mate controleren van de toegankelijkheid van doelsequenties voor tools voor genoombewerking, ook de keuzes voor DNA-reparatieroutes en andere determinanten voor DNA-bewerking beïnvloeden, waarvan de gecombineerde effecten uiteindelijk de waargenomen eindpunten van genoommodificatie bepalen. Uit eerdere onderzoeken uit de gastgroep is bijvoorbeeld gebleken dat de activiteit van designernucleasen op basis van CRISPR-Cas9-systemen en transcriptie-activatorachtige effectoren aanzienlijk wordt belemmerd door heterochromatische statussen en dat de verhouding tussen HDR en mutagene NHEJ-gebeurtenissen kan variëren, afhankelijk van de Chromatine-context.

Gezien het feit dat basiseditors en prime-editors worden gevormd door CRISPR-nickasen te fuseren met secundaire effectordomeinen (dat wil zeggen respectievelijk de-aminasen en reverse transcriptasen), wordt in **hoofdstuk 7** onderzocht of er oorzaak-gevolgassociaties bestaan tussen alternatieve chromatineconformaties en genoombewerkingseindpunten, bij gebruik van deze twee krachtige DSB-onafhankelijke genoombewerkingsplatforms. Door complementaire cellulaire systemen met functieverlies en functiewinst te implementeren, bleek dat ‘prime editing’ vaak wordt gehinderd bij heterochromatine dat wordt beïnvloed door de KRAB/KAP-1/HP1-as, alleen of samen met de DNA-methyltransferasen DNMT3A en DNMT3L. Bovendien dragen de uitgebreide delen van gRNAs die pegRNA's vormen bij aan de beperkte efficiëntie van prime-editors bij heterochromatische sequenties. Opmerkelijk is dat, in contrast met primaire editors en designernucleasen (al dan niet op basis van CRISPR), de DNA-bewerkingsactiviteit van base-editors bij gesloten heterochromatine op een doelplaats-afhankelijke manier varieert van lager tot, vaak, significant hoger dan die waargenomen bij open euchromatine. Bovendien bleek ook dat de chromatineomgeving van doelsequenties de betrouwbaarheid en zuiverheid van basisbewerkingsproducten op een gRNA-afhankelijke manier aanzienlijk kon beïnvloeden. Als gevolg hiervan vereisen de prestatie- en veiligheidsprofielen van basis- en prime-editingtechnologieën een diepgaand onderzoek om hun selectie, verdere ontwikkeling en

toepassing in specifieke contexten te begeleiden. Op basis van de gegevens gepresenteerd in **Hoofdstuk 7** kan men ook stellen dat algoritmen die zijn getraind om de activiteiten van basisbewerkings- en prime-bewerkingsreagentia te voorspellen, naast invoer op doelsequenties, zullen profiteren van de verwerking van informatie over de epigenetische context van genoemde sequenties. Tenslotte kunnen deze gegevens de ontwikkeling van combinatorische strategieën verder begeleiden, waarin gerichte epigenetische modulatoren en DSB-vrije genoombewerkingstools samenwerken om een efficiëntere en/of nauwkeurigere genetische modificatie van cellulaire (epi)genomen te bereiken.

Samenvattend verschaft dit proefschrift inzicht in hoe deze genoombewerkingstools en strategieën, die gebaseerd zijn op CRISPR-Cas9 nickases als zodanig of op hun afgeleiden voor prime editing en base editing, in menselijke cellen werken en daarmee nieuwe wegen openen voor de naadloze modificatie van cellulaire (epi)genomen. De verwachting is dat door het verder ontwikkelen en verfijnen van ‘zachte’ genoom-editing procedures die, naast de efficiëntie, rekening houden met specificiteit en nauwkeurigheid parameters, het mogelijk zal zijn om ‘genoomchirurgie’ interventies te vertalen in effectieve en veilige gen- en celtherapieën.

List of Publications

1. **Wang Q**, Capelletti S, Liu J, Janssen J and Gonçalves M. Selection-free precise gene repair using high-capacity adenovector delivery of advanced prime editing systems rescues dystrophin synthesis in DMD muscle cells. Nucleic Acids Res. (2024) Online ahead of print. DOI: 10.1093/nar/gkae057.

2. Fan C, **Wang Q**, Kuipers T, Cats D, Iyengar P, Hagenaars S, Mesker W, Devilee P, Tollenaar R, Mei H, Ten Dijke P. LncRNA LITATS1 suppresses TGF-β-induced EMT and cancer cell plasticity by potentiating TβRI degradation. EMBO J. 42:e112806 (2023).

3. **Wang Q**, Liu J, Janssen J and Gonçalves M. Precise homology-directed installation of large genomic edits in human cells with cleaving and nicking high-specificity Cas9 variants. Nucleic Acids Res. 51:3465-3484 (2023).

4. Tasca F, Brescia M, **Wang Q**, Liu J, Janssen J, Szuhai K and Gonçalves M. Large-scale genome editing based on high-capacity adenovectors and CRISPR-Cas9 nucleases rescues full-length dystrophin synthesis in DMD muscle cells. Nucleic Acids Res. 50:7761-7782 (2022).

5. Fan C, **Wang Q**, van der Zon G, Ren J, Agaser C, Slieker R, Iyengar P, Mei H and Ten Dijke P. OVOL1 inhibits breast cancer cell invasion by enhancing the degradation of TGF-β type I receptor. Signal Transduct Target Ther. 7:126 (2022).

6. **Wang Q**, Liu J, Janssen J, Tasca F, Mei H and Gonçalves M. Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. Nucleic Acids Res. 49:11986-12001 (2021).

7. **Wang Q**, Liu J, Janssen J, Le Bouteiller M, Frock R and Gonçalves M. Precise and broad scope genome editing based on high-specificity Cas9 nickases. Nucleic Acids Res. 49:1173-1198 (2021).

8. Maggio I, Zittersteijn H, **Wang Q**, Liu J, Janssen J, Ojeda I, van der Maarel SM Lankester A, Hoeben R and Gonçalves M. Integrating gene delivery and gene-editing technologies by adenoviral vector transfer of optimized CRISPR-Cas9 components. Gene Ther. 27:209-225 (2020).

9. Tasca F*, **Wang Q*** and Gonçalves M. Adenoviral vectors meet gene editing: a rising partnership for the genomic engineering of human stem cells and their progeny. Cells 9:953 (2020). * Shared first co-authorship.

10. Chen X*, Tasca F*, **Wang Q**, Liu J, Janssen J, Brescia M, Bellin M, Szuhai K, Kenrick J, Frock R and Gonçalves M. Expanding the editable genome and CRISPR-Cas9 versatility using DNA cutting-free gene targeting based on in trans paired nicking. Nucleic Acids Res. 48:974-995 (2020).

11. Xu H, Luo D, Gao X, Liu X, **Wang Q**, Sun A, Shen J, He R, Lu G, Li K, Zhang J and Xiao L. ZeoR: A Candidate for Assays Testing Enzymatic Activities and Off-Target Effects of Gene-Editing Enzymes. J. Biomed Nanotechnol. 15:662-673 (2019).

12. Yin Y*, **Wang Q***, Xiao L*, Wang F, Song Z, Zhou C, Liu X, Xing C, He N, Li K, Feng Y and Zhang J. Advances in the Engineering of the Gene Editing Enzymes and the Genomes: Understanding and Handling the Off-Target Effects of CRISPR/Cas9. J. Biomed. Nanotechnol. 14:456-476 (2018). * Shared first co-authorship.

13. **Wang Q***, Xiao L*, Zhou L, Sun W, Xing C, Li K and He N. Comparison of the Off-Target Effects Among One-Base to Three-Base Mismatched Targets of gRNA Using a Blue to White Assay. J. Nanosci Nanotechnol. 18:1594-1598 (2018). * Shared first co-authorship.

14. Xi B, Wang X, Xiao L, **Wang Q**, Ji J, Zhang L, Xue M, Liu W, Zhu R, Li K and Sun W. Gene editing effects of CRISPR/Cas on breakpoint cluster region gene targets using white-to-blue colony formation assay. Nanosci. Nanotechnol. Lett. 9:1998-2004 (2017). * Shared first co-authorship.

15. Zhou L*, **Wang Q*** and Li K. Construction of a vector with two repeats flanking CRISPR/Cas9 target for the evaluation of enzymatic activity in E.Coli. J. Nanosci. Nanotechnol. 16:12332-12336 (2016). * Shared first co-authorship.

16. Liu B, Zhou L, **Wang Q** and Li K. A mutation-sensitive switch assay to detect five clinically significant epidermal growth factor receptor mutations. Genet. Test Mol. Biomarkers 19:316-23 (2015).

Curriculum Vitae

Qian Wang was born on the 14th of March 1992 in Lianyungang, Jiangsu Province, China. From 2010 to 2014, she pursued her bachelor's degree in Pharmaceutical Engineering at Jiangsu Ocean University, China. She then started her master's study in Pharmacology at Soochow University, China, from 2014 to 2017. During her master's internship, she investigated the application of a Blue-White Colony Assay for the off-target evaluation of CRISPR-Cas9 activity under the supervision of Prof. Dr. Kai Li in the Molecular Medicine Laboratory.

In September 2017, Qian was awarded a China Scholarship Council (CSC)-Leiden University Scholarship and started her Ph.D. study 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), Leiden, the Netherlands. Throughout her Ph.D. research, she worked on; (i) investigating new strategies to improve the efficiency, specificity and fidelity of double strand break-free genome editing based on testing different donor DNA structures and engineered programmable nickases; (ii) exploring the feasibility of rescuing dystrophin synthesis in DMD muscle cells via high-capacity adenovectors encoding prime editors; and (iii) studying the impact of the chromatin context on the performance of base editors and prime editors. The work of her Ph.D. research is presented in this thesis.

From March 2022 to March 2023, Qian continued her research in Gonçalves' lab as a postdoctoral researcher, focusing on a project on gene therapy for inherited cardiomyopathies. Since March 2023 until now, she has joined the group of Prof. Dr. Peter ten Dijke in the Department of Cell and Chemical Biology at the LUMC. Her current work focuses on deciphering the functions and mechanisms of long non-coding RNAs and circular RNAs in TGF-β signaling and breast cancer progression.

Acknowledgements

Time flies. Finally, I am reaching the destination of my PhD adventure, which is definitely a challenging chapter in my life. In these years, I have tasted the joy of success and as well as experienced the frustration of failure. Glancing back, I realize that there are so many people and things worth remembering and I would like to give my sincere thankfulness to all the people who ever helped me.

First of all, I would like to express my deep gratitude to my promoter Prof. Dr. Rob C. Hoeben for his valuable input and constant support during my PhD training. I would also like to extend my heartfelt gratitude to my supervisor Dr. Manuel A. F. V. Gonçalves for his guidance, encouragement, motivation, inspiration, support and trust. He showed me the beauty of the genome editing world and let me know the importance of critical thinking and scientific responsibility. My sincere thanks go also to my co-advisors Prof. Dr. Peter ten Dijke for sharing his valuable experience and knowledge, and Dr. Harald Mikkers for his practical feedback and comments.

I am also grateful to all the members of the genome editing group. It is my pleasure to be part of you. Special thanks to Kim and Jin, who gave me enormous experimental suggestions and technical assistance on my projects. I would also like to express my appreciation to all the former and current labmates Francesca, Hidde, Li, Sabrina, Xiaoling, Igno, Marcela, Silvana and Xiaoyu for their discussions, suggestions and support.

A lot of thanks to the members of the Virus and Stem Cell Biology (VSB) Section for their fruitful discussion and kind help, and to everyone in our Cell and Chemical Biology (CCB) department. I am especially thankful to Maarten and Midory for sharing their experience and experimental materials with me, to Gerard and his wife for helping with my 'Nederlandse samenvatting', to our secretaries Julia and Pauline for their countless arrangement of meetings, to Willem for solving my computer and software problems, and to Hans for his funny jokes. I also thank my collaborator Leon (Hailiang) for his assistance in sequencing data analyses, and Annelies, Martijn, Galila, Ben and Steve for their daily help.

In addition, many thanks go to my Chinese friends for their support and the wonderful time we had together: Chao, Haijiang, Jin Ma, Jin Gan, Jiang, Jing, Jiying, Kun, Sijia, Shaozhuo, Wan, Weixin, Xueying, Xinxin, Yifan, Yongsheng and Yufeng.

Last, I would like to give my deepest thanks to my parents and brother, for their endless love and care. I am also extremely grateful to my husband Chuannan, who always cheers me up when I am upset and is willing to share all the happiness or sadness with me. His company, encouragement and support make my life full of love and fun. Finally, warm thanks go to my little boy Qi-an (Dundun) for the surprise he has brought into my life.

My PhD journey comes to an end, but memories last forever.