The Mechanism of Precise Genome Engineering in Human Cells

A DISSERTATION SUBMITTED TO THE FACULTY
OF UNIVERSITY OF MINNESOTA
BY

Yinan Kan

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR
OF PHILOSOPHY

Eric A. Hendrickson, Ph.D.
Advisor

September 2015

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Eric Hendrickson, who played a key role in the development of a Chinese boy who was upset about his academic career in the United State, into a young scientist who is excited about the infinite possibilities of the future. His encouragement woke me up with motivation every morning, and his patience helped me pull myself together in the unlucky nights.

I would like to thank my parents Shizhe Kan and Yuming Tang for their supports overseas. They cheered for all the tiny achievements that I made as if everything I did exceeded their expectations, and they never blamed me in any occasion. We could share every secret in life just like old friends.

I would also like to thank Brian Ruis for training me so patiently that I could finish my thesis research. Thank all the other past and present members in the Hendrickson Lab for making science so much fun, namely, Adam Harvey, Sehyun Oh, Euhan Lee, Elizabeth Thompson, Sarah Riman, Sherry Lin, Benjamin Akhuetie-Oni, Nicholas Mielke, Taylor Takasugi, Corey Weelock and Crystal Cheng.

Thanks to Dr. Anja Bielinsky, Dr. Alexandra Sobeck, Dr. Reuben Harris, Dr. Naoko Shima and their lab members for sharing food and thoughts in the Friday Lab Meeting. Thanks to Dr. Reuben Harris, Dr. Eric Hendrickson, Dr. Alexandra Sobeck, Dr. Do-Hyung Kim and Dr. David Largaespada for advising my scientific progress as thesis committee. Thanks to Dr. Duncan Clark for helping me and Dr. Daniel Voytas for leading the CGE conference. At last, thanks to all my peers in the Department of Biochemistry, Molecular Biology and Biophysics for cheering for my presentation in Itasca. I promise to keep the degree level of excellence throughout my career.

To Science

ABSTRACT

Genome engineering is the intentional alteration of the genetic information in living cells or organisms. Since Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR-associated 9 (CRISPR/Cas9) was repurposed for genome engineering, the “CRISPR Craze” is quickly bridging the genotype and phenotype worlds and transforming the biological, biomedical and biotechnological research. Interestingly, CRISRP/Cas9 does not perform precise genome engineering (PGE) by itself, but it only induces a targeted genomic lesion and invites the HDR pathways to introduce the desired modifications. Although PGE has a wide application in genome modification and gene therapy, the identity, property and hierarchy of the HDR pathways leading to the formation of PGE products remain obscure.

In my doctorial dissertation, I demonstrated that double-strand DNA (dsDNA) donors with a sizable central heterology preferentially utilize the double-strand break repair (DSBR) pathway in the absence and presence of chromosomal double-strand breaks (DSBs). This pathway generates long, bidirectional conversion tracts with linear distribution. In contrast, single-strand oligonucleotide (ODN) donors utilize the synthesis-dependent strand annealing (SDSA) and single-strand DNA incorporation (ssDI) pathways, respectively, depending on the strandedness of the genomic lesions and ODN donors. These pathways produce short, unidirectional and bidirectional conversion tracts with Gaussian distributions. The SDSA pathway is preferentially utilized in the presence of compound genomic lesions such as DSBs and paired nicks. In summary, this work systematically determined the identity, property and hierarchy of the HDR pathways underlying PGE with definitive molecular evidence, and provided practical guidelines for the improvement of PGE.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I:

CHAPTER II:

CHAPTER III:

LIST OF TABLES

CHAPTER II:

CHAPTER III:

LIST OF ABBREVIATIONS

CRISPR: clustered regularly interspaced short palindromic repeat

Cas9: CRISPR-associated 9

PGE: precise genome engineering, *aka*, precise genome editing

HDR: homology-directed repair, *aka*, homologous recombination

DSB: double-strand break

DSBR*: double-strand break repair, *aka*, the Szostak model

dsDNA: double-strand DNA

ODN: oligonucleotides, aka, single-strand oligonucleotides

rAAV: recombinant adeno-associated virus

HJ: Holliday junction

SDSA: synthesis-dependent strand annealing

ssDI:** single-strand DNA incorporation, *aka*, single-strand assimilation

LOH: loss-of-heterzygosity

BIR: break-induced replication

SSA: single-strand annealing

NHEJ: non-homologous end joining

C-NHEJ: classical non-homologous end joining

A-NHEJ: alternative non-homologous end joining

MMR: mismatch repair

ZFN: zinc finger nuclease

ZF: zinc finger (domain)

TALEN: transcription activator-like effector nucleases

TALE: transcription activator-like effector (module)

crRNA: CRISPR RNA

tracrRNA: trans-activating CRISPR RNA

PAM: protospacer-adjacent motif

sgRNA: short guide RNA

RNAi: RNA interference

iPS cells: induced pluripotent stem cells

PCR: polymerase chain reaction

HPRT: X-linked hypoxanthine phosphoribosyltransferase

Neo^{R}: neomycin-resistance gene

PIGA: phosphatidylinositol glycan anchor biosynthesis, class A

SNP: single nucleotide polymorphism

nCas9*:** the nickase version of Cas9

S**:** sense

AS**:** antisense

IR: ionizing radiation

D-loop: displacement loop, *aka*, D-shaped loop

*In this dissertation, **DSBR** refers to the specific Szostak model of HDR, whereas **DSB repair** refers to the general double-strand break repair processes including HDR and NHEJ.

ssDI is coined as an alternative designation of single-strand assimilation in order to distinguish it from single-strand annealing (**SSA**).

***In this dissertation, **nCas9** specifically refers to the Cas9 D10A variant that nicks the opposite strand of the sgRNA.

****These abbreviations only apply to **Chapter III**.

LIST OF PROJECTS AND PUBLICATIONS

1. **Kan Y**, Ruis B, Lin S and Hendrickson EA (2014). The mechanism of PGE in human somatic cells. ***PLoS Genetics***, 10:e1004251.

   A modified version of this publication is included in **Chapter II**. Conceived and designed the experiments: **YK** EAH. Performed the experiments: **YK** BR SL. Analyzed the data: **YK** EAH. Wrote the paper: YK EAH.

2. **Kan Y**, Mielke N and Hendrickson EA (2015). The mechanisms of precise genome editing using oligonucleotide donors. **In review.**

   A reprint of this publication is included in **Chapter III**. Conceived and designed the experiments: **YK** EAH. Performed the experiments: **YK** NM. Analyzed the data: **YK** EAH. Wrote the manuscript: YK EAH.

3. **Kan Y**, Akhuetie-Oni B, Mielke N, Takasugi T and Hendrickson EA (2015). The kinetics of precise genome editing as a function of homology arm length. **In preparation**.

   This manuscript is not included in this dissertation. Conceived and designed the experiments: **YK** EAH. Performed the experiments: **YK** AB NM TT. Analyzed the data: **YK** EAH.

4. **Kan Y**, Riman S and Hendrickson EA (2015). The single-strand annealing activity of human RAD52 is involved, but not required, in all the homology-directed repair pathways. **In preparation.**

This manuscript is not included in this dissertation. Conceived and designed the experiments: **YK** EAH. Performed the experiments: **YK** SR. Analyzed the data: **YK** EAH. Wrote the manuscript: YK EAH.

5. Fattah FJ, Kweon J, Wang Y, Lee EH, **Kan Y**, Lichter N, Weisensel N and Hendrickson EA (2014). A role for XLF in DNA repair and recombination in human somatic cells. ***DNA Repair (Amst)****, 15:39-53.*

   This publication is not included in this dissertation.

6. Schumacher AJ, Mohni KN, **Kan Y**, Hendrickson EA, Stark JM and Weller SK (2012). The HSV-1 exonuclease, UL12, stimulates recombination by a single strand annealing mechanism. ***PLoS Pathogens***, 8:e1002862.

   This publication is not included in this dissertation.

7. **Kan Y**, Hendrickson EA and Bürckstümmer T et al. Haploid human genetics enables the screening of suppressors in the homology-directed repair pathways. ***Work in Progress.***

   This study is being performed in collaboration with Horizon Discovery. Conceived and designed the experiments: **YK** EAH TB. Performed the experiments: TB et al.

8. **Kan Y.** Streamlined production of dsDNA donors and bi-allelic edited human cells. ***Unpublished Protocol.***

   This protocol is confidential and not included in this dissertation. However, it was used to establish numerous collaborations that may result in publications.

CHAPTER I: INTRODUCTION

Background

All the genetic information (*i.e.* the genome) of a cellular organism is stored in the form of DNA [1]. Human somatic cells carry about 3.2 billion base pairs of DNA arranged in 23 pairs of enormously long molecules called chromosomes. These chromosomes contain a total of approximately 20,000 molecular units (*i.e.* genes), which encode proteins and catalytic RNAs that substantiate the vast majority of cellular functions, such as metabolism, signaling, maintenance and proliferation. Furthermore, the functions of about 3.7×10^{13} cells in our body [2] are collaborated to support the daily activities of our living including perception, cogitation, motion and reading this dissertation.

The faithful maintenance and transmission of genetic information is crucial to all forms of cellular and organismal lives. However, genomic lesions can arise from errors of endogenous DNA replication and exposure to exogenous DNA damaging agents. DSBs are the most lethal forms of genomic lesions to living cells. DSBs occur about 10 times per cell per day, and a single unrepaired DSB may lead to cell death [3]. Moreover, the improper repair of DSBs may lead to mutagenesis and tumorigenesis [4, 5]. Eukaryotes have developed two major pathways of DSB repair: HDR and non-homologous end joining (NHEJ). HDR precisely repairs DSBs using the redundancy of genetic information on the homology donors [6, 7]; whereas NHEJ efficiently joins DNA ends together after minimal necessary processing [8-10].

For the first time, genome engineering allows human beings to intentionally manipulate the hereditary materials of living organisms that are otherwise only accessible to natural evolution. Although cumbersome in the beginning, the development of artificial meganucleases, especially CRISPR/Cas9, dramatically

accelerated the process of genome manipulation. Nowadays, plants and animals with desired genotypes can be produced in laboratories within one life cycle of these model organisms, which greatly facilitate the deciphering of the astronomic amount of genetic information generated by the genome sequencing projects. In clinics, gene therapies are in various stages of human trials [11], which hold the promise of permanently treating inherited diseases and incurable viruses. Interestingly, meganucleases do not perform genome engineering by themselves, but they induce targeted genomic lesions and facilitate either high-efficiency gene disruption via NHEJ or, in the presence of homology donors, PGE via the HDR pathways [12]. Although PGE generates precisely predictable outcomes, its efficiency largely depends on the HDR machinery and cell cycle status of the specific organisms. Importantly, the mechanisms and characteristics of the HDR pathways leading to PGE are not well understood in human cells.

In this dissertation, **Chapter I** contains a comprehensive introduction of the DNA DSB repair pathways and the development of PGE technologies. In **Chapter II** and **chapter III**, the molecular mechanisms of PGE in human cells using dsDNA and ODN donors are dissected with definitive experimental evidence. In **Chapter IV**, the significance of this study and the prospects of PGE are discussed.

DNA Double-Strand Break Repair

"We totally missed the possible role of … (DNA) repair although … I later came to realise that DNA is so precious that probably many distinct repair mechanisms would exist."

Francis Crick wrote in *Nature*, 1974 [13]

Homology-directed repair

HDR, *aka* homologous recombination, is the exchange of genetic information between similar or identical sequences. It is an accurate way of repairing DNA DSBs using the redundancy of genetic information on the homology donors, such as the undamaged sister chromatids or homologous chromosomes. In budding yeast *Saccharomyces cerevisiae*, HDR repairs the bulk of ionizing radiation (IR)-induced DSB, whereas NHEJ simply fails to do so for the lack of end processing enzymes [14]. In mammalian cells, in contrast, HDR plays a minor role in DSB repair due to the robust competition from non-homologous end joining (NHEJ) [15]. HDR repairs a fraction of DSBs mainly in the S and, to a lesser extent, in the G2/M phases of the cell cycle [16]. It is also involved in repairing complex genomic lesions such as Holliday junctions (HJs), interstrand crosslinks (ICLs) and stalled replication forks (SRFs). Besides DNA repair and replication, HDR is also required for the meiotic segregation and telomere maintenance [6]. Furthermore, the dysfunction of HDR is associated with cancer pre-deposition [4, 5] and aging [17].

The genetic players in the HDR pathways can be easily identified in yeast as mutants sensitive to IR [18]. Because the yeast *RAD52* is required in all the HDR

pathways, most of these HDR mutants fall into the *RAD52* epistasis group [19, 20], which currently includes: *RAD50*, *RAD51*, *RAD52*, *RAD53*, *RAD54*, *RAD55*, *RAD56*, *RAD57*, *RAD59*, *MRE11*, *XRS*2 and *RDH54/TID1* [6, 7, 21]. The functions of these genes are generally well conserved in the mammalian HDR, despite the existence of multiple homologs of *RAD51*, *RAD54* and *RAD55-77*, and the emergence of *BRCA1* and *BRCA2* that partially replace the function of *RAD52*. However, the classification of HDR genes into the *RAD52* epistasis group masks the subdivision of this complicated molecular gymnastics into at least six different pathways [21] as elaborated below:

The double-strand break repair model

The DSBR (aka, Szostak) model was first elaborated by Szostak JW. et al. [22, 23] based on the conceptual contributions from Holliday R [24], Meselson MS. and Radding CM. [25]. In this model, as is illustrated in **Figure 1A,** after HDR is initiated by DSB formation, the 5' ends of the DSB are processed by nucleolytic resection. The resection gives rise to free single-strand 3' ends, and one of them invades a double-strand homology donor during homology search. DNA synthesis then extends the invading 3' end using the donor sequence as a template, and the displacement synthesis leads to the formation of a displacement loop (D-loop). Importantly, the second end of the DSB is captured and annealed to the extending D-loop in this DSBR model, leading to the formation of two HJs. The resolution of the double HJs by resolvases generates both crossover and non-crossover products.

The DSBR model is characterized by the generation of equal amounts of crossover and non-crossover products. Moreover, one of the non-crossover products contains the signature bidirectional conversion tract in respect to the DSB. In addition, as is

elaborated in **Chapter II**, this dissertation also revealed a novel molecular feature of the DSBR pathway, which is long-tract gene conversion.

Holliday junction dissolution

Although the crossovers generated by the DSBR pathway are required for the proper segregation of chromosomes during meiosis, they may cause deleterious mutations and are thus suppressed in mitotic cells. The HJ dissolution pathway emerges as an alternative way of solving double HJ structures, in which the HJs cancel out each other as a result of inward branch migrations catalyzed by a HJ dissolution complex composed of BLM, TopoIIIα, RMI1 and RMI2 [26-28]. As is illustrated in **Figure 1B**, this pathway is identical to DSBR till the second strand capture step. Then branch migration occurs and, instead of resolved by the endonuclease cleavage, the HJs simply cancel out each other in the help of the BLM complex.

Because the invading strand eventually anneals back with its original partner, gene conversion stems from the unidirectional DNA synthesis using the homology donor as a template. As a result, this pathway only generates non-crossover products. Since a large D-loop can be stabilized by second strand capture, we reason that HJ dissolution is also a long-tract gene conversion pathway similar to DSBR.

Synthesis-dependent strand annealing

The DSBR model was originally proposed according to the observations of the meiotic recombination in yeast. However, mitotic recombination is infrequently associated with crossovers in yeast, fruit flies and mammalian cells [29-32]. In addition, DSB-induced gene conversion in mouse and human cells predominantly produces short

unidirectional conversion tracts [33]. These observations leads to the proposition of SDSA as the major pathway of HDR-mediated DSB repair in mitotic cells [6, 7, 21].

The SDSA pathway, as is illustrated in **Figure 1C**, is similar to the DSBR model till displacement synthesis. After the D-loop formation, in contrast, the synapse collapse before the second end capture, and the invading strand anneals back with the second end of the homology recipient on the other side of the DSB. Similar to the HJ dissolution pathway, the donor sequence is converted unidirectionally as a result of DNA synthesis, and this pathway produces only non-crossover products. However, as is demonstrated in **Chapter III** and another publication [33], SDSA is a short-tract gene conversion pathway that may result in the loss-of-heterozygosity (LOH) in DSB repair.

Break-induced replication

Early studies of DSB-induced gene conversion revealed a unidirectional Bi-Gaussian distribution of the conversion tracts in yeast, mouse and human cells [33, 34]. The generation of ultra long conversion tracts suggested the existence of an alternative HDR pathway that was later proved to be BIR [35, 36]. In this pathway, as is illustrated in **Figure 1D**, the second end of the DSB is either lost or non-existent (such as collapsed replication forks or telomeric DSBs). Thus the invading end has to perform long-tract displacement synthesis until it travels through the rest of the template chromosome or encounters a converging replication fork [36]. Then a complementary single-strand DNA primer fills in to initiate the reverse synthesis similar, but not exactly the same, to the lagging strand synthesis in a replication fork.

Because the entire chromosome on the other side of the DSB has to be reconstituted via DNA synthesis, BIR can lead to long-tract LOH in the chromosomal

scale. Also, the template switching process can be highly mutagenic. As a result, this pathway is usually suppressed in the presence of the second end [35].

Single-strand annealing

Occasionally, DSBs occur in between direct repeats such as microsatellites sequences, which allow the self-templated repair via a homology-directed deletion process called SSA. In this pathway, as is illustrated in **Figure 1E**, the DNA ends are resected to reveal complementary single-strand 3' overhangs that are subsequently annealed together to bridge the DSBs. This pathway results in the deletion of the non-homology sequence in between the direct repeats as well as one copy of the repeats. The strand annealing process requires the activity of RAD52 in yeast, but is independent of strand invasion and HJ formations in the DSBR pathway [7, 37].

Single-strand DNA incorporation

Unlike all the HDR pathways mentioned above, ssDI (aka, single-strand assimilation) is not generally regarded as an endogenous DNA repair process. However, it is widely used in the PGE field to describe the incorporation of exogenous single-strand DNA, including ODNs [38-40] and recombinant adeno-associated virus (rAAV) [41-43] donors, into genomic lesions such as DSBs. Although this pathway occurs at a relatively low frequency, its existence was confirmed by the physical incorporation of ODNs into the targeted chromosome locus in the absence of exogenous genomic lesions [38]. Interestingly, a variation of this pathway can also occur via a single-strand RNA intermediate, in which the transcript RNA bridges one strand of the DSB without being physically incorporated into the genome [44].

In the ssDI pathway, as is illustrated in **Figure 1F**, a single-strand DNA or RNA donor is assimilated to a single-strand gap via homology base-paring with the complementary strand. After the displacement of the flanking homology sequences via branch migration like activities, the DNA flaps are cleaved by endonucleases and part of the donor sequence is physically ligated into the chromosome to bridge the single-strand gap. If the gap is generated by DSBs, the resulting nick on the other strand is further repaired by standard single-strand break repair. The ssDI pathway is characterized by the bidirectional gene conversion and the physical incorporation of the donor sequence into chromosomes [38]. It was also demonstrates that ssDI is a short-tract gene conversion pathway in **Chapter III** of this dissertation.

Non-homologous end joining

NHEJ is the major pathway of repairing IR-induced DSBs in human cells. NHEJ has robust DSB repair activity in all phases of the cell cycle, which increases as cells progress from the G1 to the G2/M phases [16]. NHEJ is also required for the repair of programmed DSBs during V(D)J recombination in lymphocytes maturation [15, 45, 46]. In addition, evidence suggests that NHEJ is the major source of chromosomal translocations during cancer development [47, 48].

Unlike HDR, the NHEJ pathways do not require extended homology in repairing DSBs. Instead, the NHEJ proteins bind to the broken DNA ends, and ligate them together after minimal necessary processing. Depending on the extent or arguably the property of the microhomology being utilized, the NHEJ pathways can be divided into canonical NHEJ (C-NHEJ) and alternative NHEJ (A-NHEJ). These pathways seem to utilize different panels of protein players, although the specific molecular distinctions remain controversial [9].

Precise Genome Engineering

> *"(CRISPR/Cas9) is really a triumph of basic science and in many ways it's better than RNA interference. It's a tremendous breakthrough with huge implications for molecular genetics. It's a real game-changer."*
>
> Craig Mello told *The Independence*, 2013 [49]

A brief history of genome engineering

Genome engineering is the intentional alteration of a genetic locus in living cells or organisms. According to the purpose of engineering, it can be classified into genetic knock-out and genetic knock-in. Genetic knock-out is the targeted inactivation of an endogenous gene, whereas genetic knock-in is the introduction of a desired modification into a designated genetic locus. Both of these approaches take advantage of the endogenous DNA DSB repair processes. According to the specific DNA repair pathway being utilized, genome engineering can be divided into the NHEJ-mediated gene disruption, and the HDR-mediated PGE.

PGE precisely convert the genetic information from exogenous homology donors to a targeted locus on the chromosome via the HDR pathways. Notably, this approach can be applied to both genetic knock-out and genetic knock-in. Since PGE generates precisely predictable outcomes in the conversion products, it represents the more desirable pathway of genome engineering and holds promise for correction-based gene therapy. However, because of the limited HDR activity in the somatic cells of higher eukaryotes [8, 16], the efficiency of spontaneous PGE (i.e., PGE in the absence of

targeted genomic lesions) is relatively low and highly variable depending on the cell cycle status and the specific research model being used [50].

Interestingly, the efficiency of PGE can be stimulated by more than 1,000 fold in the presence of targeted genomic lesions near the target locus, such as DSBs created by meganucleases [51-53]. This observation started a new era of PGE using artificial meganucleases (*aka*, designer or programmable meganucleases). Zinc finger nucleases (ZFNs) [54], transcription activator-like effector nucleases (TALENs) [55] and CRISPR/Cas9 [56-60] were sequentially developed such that targeted genomic lesions could be introduced with increasing ease. Unlike natural meganucleases, the artificial meganucleases can be engineered to bind to, in principle, any designated sequences in the genome, which obviates the need of knocking in their recognition sites near the target locus, ironically via PGE itself. After DNA recognition, the meganucleases create genomic lesions and facilitate high-efficiency gene disruption via NHEJ, or PGE via the HDR pathways in the presence of exogenous homology donors [12]. In general, these homology donors can be classified into dsDNA (including plasmid vectors and viruses with dsDNA intermediates), rAAV and ODNs.

Homology donors

Double-strand DNA donors

To the best of my knowledge, the first PGE experiment in eukaryotic cells was carried out by Hinnen A. *et al* in 1987 [61]. As a corollary discovery of the Calcium chloride transformation method in yeast [61], it was discovered that a large fraction of the transformed plasmids were precisely targeted into the homologous chromosomal locus when circular plasmids with homology region were utilized [62]. However, this PGE technology did not become popular until Rothstein R. and collaborators

demonstrated that its frequency can be dramatically increased by linearizing the plasmid at both ends of the homology regions [63]. These findings also led to the development of the DSBR model of homologous recombination [22, 23], which was later proved to be the major mechanism of mitotic HDR and PGE in yeast [64]. The idea of PGE using linear dsDNA donors via the spontaneous "ends-out" recombination was later extended to the genetic knock-outs and knock-ins in mammalian cells and won Smithies O. and Capecchi MR. the Nobel Prize in 2007 [65, 66]. In a standard dsDNA donor, the knock-in mutation is flanked by long homology arms ranging from 0.8-12 kb.

Recombinant adeno-associated virus donors

Although spontaneous PGE using dsDNA donors works relatively well in yeasts, flies and murine embryonic stem cells, its efficiency drops to as low as 10^{-6} in human cells due to the weak HDR activity and the robust competing pathway of C-NHEJ. Isolating the rare PGE products thus requires the laborious screening of thousands of individual subclones even in the presence of drug selection [67-69]. Importantly, Russell DW. and coworkers reported that the efficiency of spontaneous PGE could be increased up to 3 orders of magnitude by the use of rAAV donors [70]. The use of rAAV donors along with the development of advanced promoter trapping strategies collaboratively increased the efficiency of PGE to a comfortable range in human cells [71].

Adeno-associated virus is a single-strand DNA virus. Its entire 4.7 kb genome is composed of two viral genes flanked by structural elements called inverted terminal repeats. Because of its low pathogenicity and immunogenicity and broad tropism in human and mouse cells, rAAV was widely used for delivering exogenous DNA into human cells for transfection and gene therapy. However, it was not known why rAAV also works so well as a PGE donor. Some papers suggested that this was because,

unlike dsDNA donors, rAAV-mediated PGE could utilize the ssDI pathway [41-43]. Nevertheless, rAAV can become double-stranded during replication, and it is not known whether PGE is initiated by the single-stranded or double-stranded viral intermediates.

Single-strand oligonucleotide donors

ODNs cannot initiate efficient PGE in the absence of meganucleases because they have to engage endogenously occurring genomic lesions. It was proposed that ODNs could be incorporated into the gaps generated by nucleotide excision repair [72, 73] or the lagging strand synthesis during DNA replication [74-76], although the specific mechanisms are still disputed [77, 78]. Notably, both of these models involve the physical assimilation of the ODNs into the genome via an ssDI-like mechanism.

Interestingly, ODNs are much more potent PGE donors in the presence of targeted genomic lesions induced by meganucleases [79], probably due to the use of an alternative pathway other than ssDI. Consistent with this notion, it was demonstrated that the ODN donors were physically incorporated into the genome in the absence, but not presence, of meganucleases [38, 79]. Because ODN donors can be easily chemically synthesized, they are widely used in conjunction with ZFNs, TALENs and CRISPR/Cas9. Unlike dsDNA donors, ODN can also mediate efficient PGE in the presence of targeted single-strand nicks. It was proposed that nick-induced PGE using ODN donors utilizes certain forms of the SDSA and the ssDI pathways depending on the relative strandedness of the ODN donors and the nick [40, 80]. However, the exact mechanisms of ODN-mediated PGE remained elusive.

Artificial meganucleases

Meganucleases are endonucleases with a large recognition site, which, in principle, does not occur naturally in a given genome. They can be used to introduce targeted lesions (such as DSBs) in the genome and dramatically enhance the efficiency of PGE using all kinds of donors [51, 52, 79]. Natural meganucleases (aka, homing endonucleases) such as I-SceI and I-AniI are the first generation of meganucleases used in PGE [81, 82]. Due to the extremely low programmability, ironically, their recognition sites have to be pre-engineered into the genome before they can be used to induce PGE. Thus the application of natural meganucleases in PGE is mainly limited to proof-of-principle experiments and mechanistic studies.

Zinc finger nucleases

To increase the programmability of meganucleases, Chandrasegaran and colleagues fused the catalytic domain of a Type IIs restriction enzyme FokI to the zinc finger DNA recognition domains (ZFs), and demonstrated that the specificity of the FokI catalytic domain could be completely dictated by the ZF arrays [54]. ZFs are modular DNA binding domains originally identified in the sequence-specific transcription factors [83] of eukaryotic cells. Each ZF binds to approximately 3 bp of double-strand DNA, and multiple modules can be combined to recognize a wide range of unique target sequences in the genome. By fusing the designated ZF arrays to the catalytic domain of FokI, the resultant ZFN can be programmed to cut almost any specific region in the genome. Also, the individual FokI monomers can be targeted to the adjacent regions of the genome by different ZF arrays to increase the specificity of the resultant ZFNs [84].

However, because the binding affinity and specificity of ZFNs are dependent on both the individual ZFs and their context, the activity and specificity of ZFNs are

generally lower than the later generations of artificial meganucleases [12, 85]. Also, since each individual ZF recognizes a tri-nucleotide codon, a large ZF library has to be kept in order to routinely create the custom ZF arrays and target random sequences in the genome. Thus the application of ZFNs is mainly limited to experienced laboratories and companies, and the cost remains high despite the effort of streamlined production.

Transcription activator-like effector nucleases

Plant pathogens such as *Xanthomonas* encode transcription activator-like effectors (TALEs) that can specifically up-regulate the host genes to promote infectivity. The DNA binding domains of TALEs are composed of tandem repeats of ~34 amino acids, and each module of the repeats recognizes a single base pair of double-strand DNA [86]. The DNA binding specificity of these modules are completely determined by 2 variable amino acid residues in the 12-13 position, which is named repeat variable diresidues (RVDs), whereas the backbone sequences of these modules are conserved. Interestingly, the TALE modules can be reprogrammed to bind almost any designated sequences in the genome, although certain restrictions may apply [12].

When a designer TALE domain is fused to the catalytic domain of FokI, it can direct the non-specific nuclease to cut designated sequences in the genome [55]. Because the DNA binding modules have a 1:1 correspondence with the target nucleotide sequences, the generation of custom designed TALENs requires a much smaller library of DNA binding modules than ZFNs. Also, the Golden Gate ligation has been adapted to the cloning of DNA binding modules into TALEN, which dramatically reduce the time cost of TALEN assembly [87]. As a result, TALENs are widely used in medium-sized laboratories and companies for genome engineering.

Clustered regularly interspaced short palindromic repeat/CRISPR-associated 9

Although the assembly of ZFNs and TALENs can be somehow streamlined, it still requires the *de novo* assembly of large DNA binding domains. The CRISPR/Cas9 system circumvents this requirement because the specificity of Cas9 is completely dictated by a 20 bp sequence on the CRISPR RNA (crRNA) complementary to the protospacer. This feature of CRISPR/Cas9 obviates the requirement of protein engineering and leads to the era of RNA-guided genome engineering.

CRISPR/Cas9 was originally identified as an adaptive immune system in bacteria and archaea [88]. These microbes capture short DNA fragments from invading (defective) phages and integrate them as protospacers into the CRISPR locus into their genomes as immunological memory [89]. When these microbes are intruded by phages containing similar protospacers again, the invading genomes can be specifically cleaved by the Cas9 complex in an RNA-dependent manner. The Type II CRISPR system is most frequently used in genome engineering because it contains only one protein (Cas9) and two RNA components (crRNA and trans-activating crRNA, *i.e.*, tracrRNA) [56]. Cas9 first binds to the crRNA and tracrRNA before it is recruited onto the genome depending on the density of the protospacer-adjacent motif (PAM). The PAM is a short sequence motif (*e.g.* NGG for the Cas9 from *S. pyogenes*) located in the vicinity of the protospacer in the invading genomes but not the CRISPR arrays of the host genomes. It plays an important role in distinguishing between “self” and “non-self” during Cas9-mediated cleavage [90]. After the recruitment of the Cas9 complex, it opens up the DNA double-helix and establishes base-paring between the crRNA and the putative protospacer sequence, resulting in the formation of a D-loop. Mismatches between the crRNA and the seed sequence (the 8-13 bp PAM proximal sequence) typically leads to

D-loop rejection, although mismatches in the PAM distal region may be better tolerated [91, 92]. When the D-loop is allowed to mature, however, it triggers a conformational change of Cas9 to cleave the protospacer sequence approximately 3 bp upstream of the PAM [93]. Interestingly, these two RNA components of the Cas9 complex can be combined into a short guide RNA (sgRNA), which further simplifies the CRISPR/Cas9 system [56]. The Cas9/sgRNA complex can be targeted to anywhere in the genome by reprogramming the 20 bp protospacer on the RNA component, obviating the need of complicated protein engineering. As a result, the sheer simplicity and user-friendliness of the CRSIRP/Cas9 system makes it a game-changer of genome engineering and gives birth to the "CRISPR Craze".

The "CRISPR Craze"

Since CRISPR/Cas9 was repurposed for genome engineering in the beginning of 2013 [57-60], more than 1500 papers have been published in the last 2.5 years, documenting the application of this technology in biological, biomedical and biotechnological studies [94-99]. As for genome engineering, the "CRISPR speed" is quickly bridging the genotype and phenotype worlds and facilitating the generation of genetically modified research models within a single life cycle of the specific host organisms [100-103]. In addition, the facile adaptation of this technology via the codon optimization of Cas9 is quickly increasing the list of genome-tractable organisms [12]. Beside genome engineering, a wide range of tools have been created for targeted transcriptional activation/suppression [104, 105], epigenetic regulations [106], chromosomal gymnastics [94] and live imaging [107] by fusing a catalytically dead Cas9 to other functional domains. These tools are widely used in genomic screening [108, 109], genetic testing [110], metabolic engineering [95, 111], synthetic biology [112, 113],

pathogen elimination [114, 115], gene drive [115, 116], drug identification [109, 117] and human gene therapy [118, 119]. As for now, the list of publications with CRISPR/Cas9 is growing in an exponential manner, as the “CRISPR Craze” continues. Scientists are already starting to compare CRISPR/Cas9 to the most influential fundamental technological platforms such as RNA interference (RNAi), induced pluripotent stem cells (iPS cells), polymerase chain reaction (PCR)…, and probably even beyond [49, 97, 99].

Figures and Legends

Figure 1. The molecular mechanism of the HDR pathways

(A-F) Schematic illustration of the mechanism of DSBR (A), HJ dissolution (B), SDSA (C), BIR (D), SSA (E) and ssDI (F). Schematic elements are colored as follows: recipient genomic DNA, black; genomic lesions, orange; homology donor, red; DNA synthesis, dashed lines; HJ resolvase cleavages, white arrows; branch migration, grey arrows; exonuclease resection, yellow pacmen^{TM}; flap-/mega- endonuclease cleavage, yellow lightning bolts.

Figure 1A-B

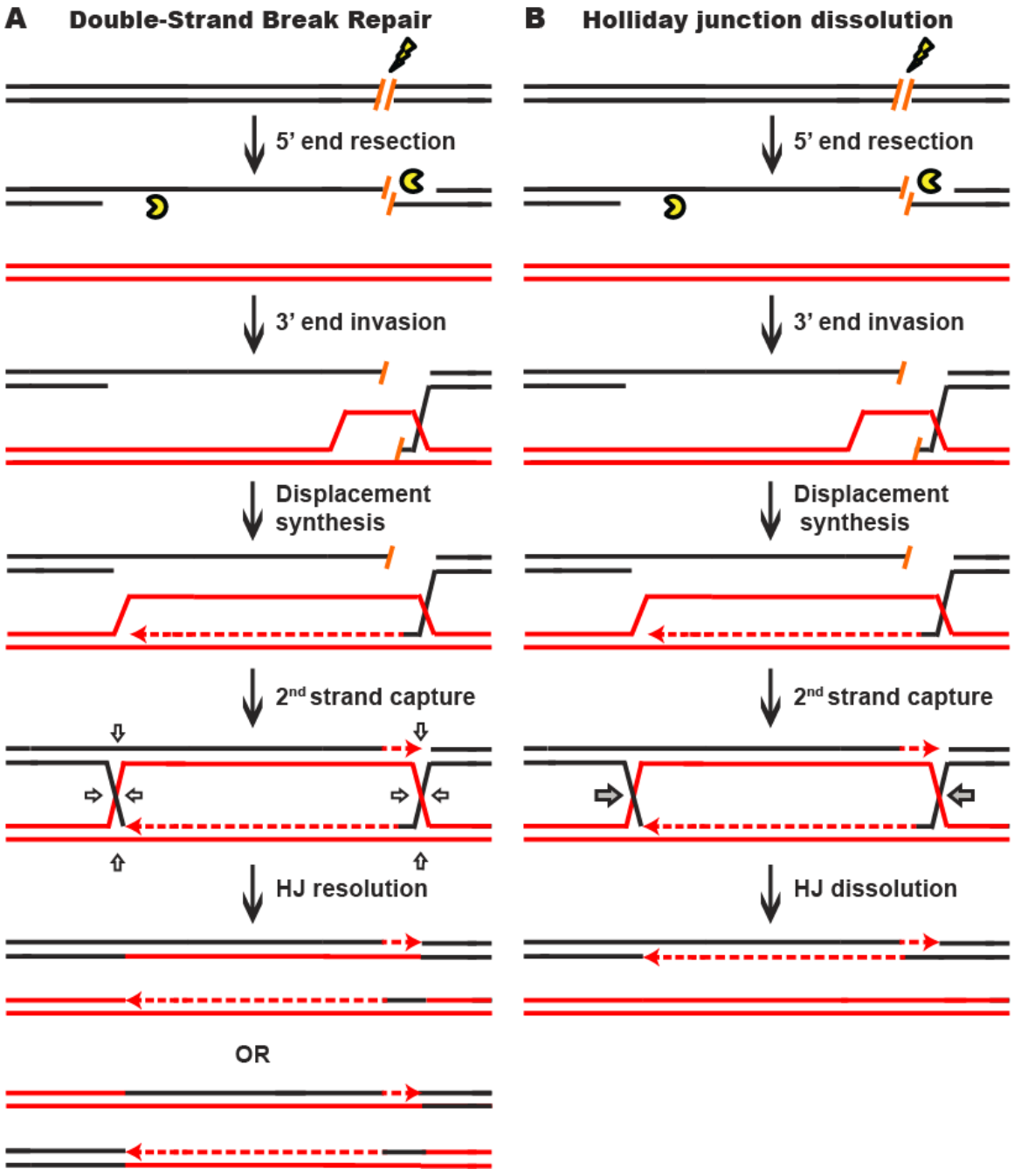

A
Double-Strand Break Repair
5' end resection
3' end invasion
Displacement synthesis
2nd strand capture
HJ resolution
OR
B
Holliday junction dissolution
5' end resection
3' end invasion
Displacement synthesis
2nd strand capture
HJ dissolution

Figure 1C-D

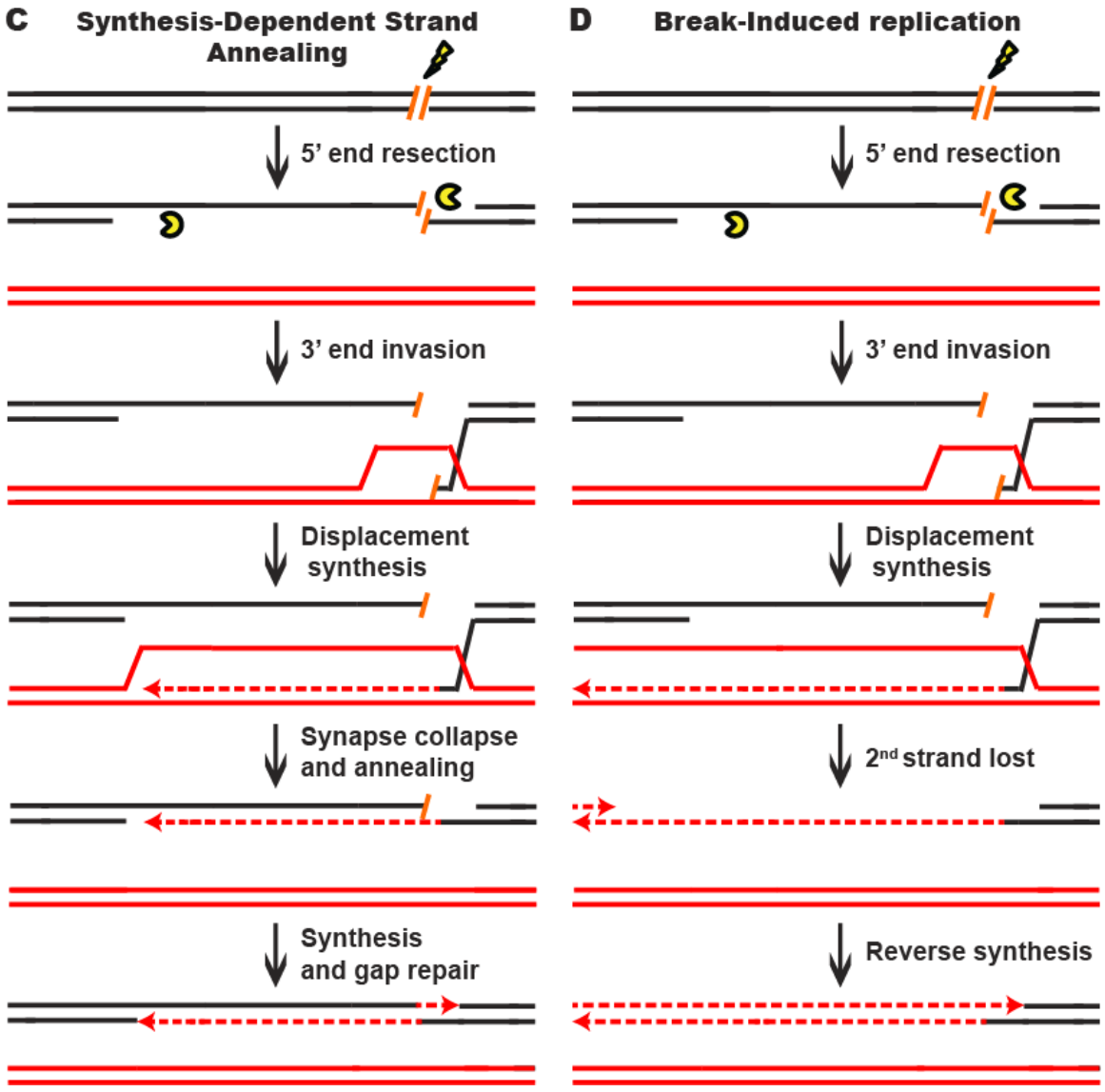
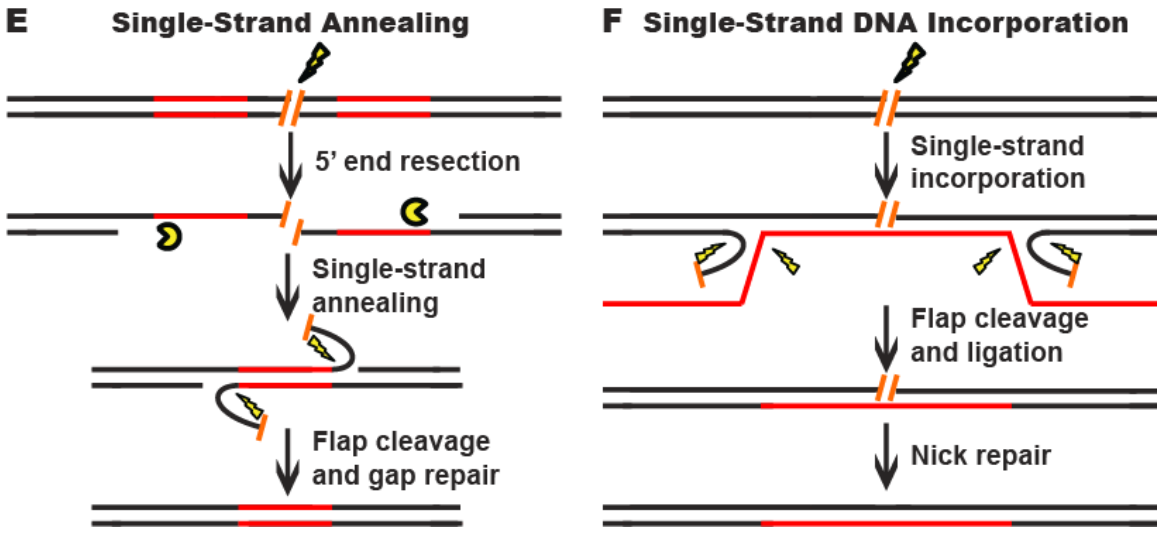

C
Synthesis-Dependent Strand Annealing
5' end resection
3' end invasion
Displacement synthesis
Synapse collapse and annealing
Synthesis and gap repair
D
Break-Induced replication
5' end resection
3' end invasion
Displacement synthesis
2nd strand lost
Reverse synthesis

Figure 1E-F


E
Single-Strand Annealing
5' end resection
Single-strand annealing
Flap cleavage and gap repair
F
Single-Strand DNA Incorporation
Single-strand incorporation
Flap cleavage and ligation
Nick repair

CHAPTER II: THE MECHANISM OF PRECISE GENOME ENGINEERING USING dsDNA DONORS

Summary

PGE in human somatic cells is of importance because it can be used to either delineate the loss-of-function phenotype of a gene or introduce designer modifications to the genome. Both of these outcomes require a form of DNA DSB repair known as HDR. The mechanism of HDR leading to PGE, however, is not well understood in human cells.

Here, we surprisingly demonstrate that the normally single-stranded rAAV performs PGE via double-stranded intermediates, which are mechanistically indistinguishable from dsDNA donors with central heterology. Using these dsDNA donors, the canonical two-end, ends-out HDR intermediate in yeast is also valid for human cells. Furthermore, the resolution step of this intermediate occurs via the classic DSBR model of HDR, whereas SDSA and Holliday junction (HJ) dissolution are, at best, minor pathways. Moreover, this DSBR pathway produces long bidirectional conversion tracts with linear distribution, which suggests that the positions of HJ resolution are evenly distributed along the homology arms. Most unexpectedly, we demonstrate that when a meganuclease is used to introduce a chromosomal DSB at the target site, the mechanism of PGE is inverted such that the chromosome becomes the “attacker” instead of the “attackee”. These observations significantly advance our understanding of HDR and PGE using dsDNA donors in human cells, and should readily lend themselves to developing improvements to existing methodologies

Introduction

PGE is the intentionally alteration of the genetic information in living cells or organisms [120]. This technology has at least two applications of significant importance. One application is the inactivation of genes ("knockouts"), a process in which the two wild-type alleles of a gene are disrupted to determine the loss-of-function phenotype associated with that particular gene. The second application is the introduction of designer modifications to the genome ("knock-ins"), including the clinically-relevant process of gene therapy, which in a strict sense, involves correcting a preexisting mutated allele of a gene back to wild-type to alleviate the pathological phenotype associated with the mutation. Importantly, although these two processes are conceptually reciprocal opposites of each other, they are mechanistically identical because both require a form of DNA DSB repair termed HDR.

During HDR, as elaborated predominately in yeast [20], the ends of the invading double-stranded DNA are resected to yield 3'-ssDNA overhangs [14], which, in turn, are substrates for Rad51. Rad51 is a strand exchange protein [121], which facilitates the base pairing of the invading strand with its homologous chromosomal donor. After second strand capture, a recombination intermediate is generated with two HJs that is identical to the intermediate of dsDNA-mediated PGE that has been well-defined in yeast [20, 64, 122, 123]. Resolution of this intermediate requires different combinations of polymerases, helicases, nucleases and ligases that result in distinct recombination products. Importantly, human cells express all of the HDR genes needed to carry out PGE [120]. However, because of the robust competing pathway of DSB repair known as NHEJ [15], PGE events occur rarely in mammals [124-126]. Indeed, despite valiant efforts — in particular by the Baker laboratory [125, 127, 128] — the low targeting

efficiency of dsDNA donors has prohibited a systematic characterization of recombination intermediates in mammalian cells. To gain better insight into the mechanism of human PGE it is crucial to establish a more vigorous PGE system.

Russell and coworkers have demonstrated that rAAV can target the human genome with frequencies up to 1% [70] (**Figure S1**), which is 3 to 4 orders of magnitude higher than dsDNA-mediated PGE. rAAV has subsequently become a powerful tool to engineer knockout and knock-in mutations in the human genome [120, 129]. Despite its utility, the mechanism of rAAV integration remains elusive although it is clear that the recombinant virus, which encodes no viral proteins, must utilize host DSB repair pathways for its integration. Interestingly, since only single-stranded genomes can be packaged into virions (**Figure S1**), many reviews [41-43] have postulated that rAAV-mediated PGE utilizes the ssDI pathway.

Here we systematically analyzed the molecular features of PGE intermediates using rAAV and dsDNA donors. In contrast to popular belief, we demonstrate that rAAV with central heterology predominantly utilizes the DSBR model of HDR [130] with double-stranded viral DNA utilized as a substrate. Specifically, we analyzed the retention of single nucleotide polymorphisms (SNPs) — markers that allowed us to distinguish donor from recipient DNA — during PGE and random integration. We show that, in contrast to lower eukaryotes and murine embryonic stem cells [131-134], the positions of HJ resolution are evenly distributed along the homology arms of the donors (**Figure S2**) in two independent human cell lines. In addition, we demonstrate that rAAV-mediated PGE events are mechanistically distinguishable from random integration events. Most unexpectedly, we observed that in the presence of chromosomal DSBs rAAV switches to a chromosome-initiated, ends-in recombination mode (**Figure S3)**, which greatly augments the PGE process. A detailed analysis of the intermediates of

the ends-in recombination reaction revealed that the DSBR model is preferred over SDSA and HJ dissolution in DSB-induced PGE using rAAV donors with central heterology. These observations greatly expand our understanding PGE and its underlying HDR mechanism in human cells.

Results

The HPRT targeting system

The X-linked *hypoxanthine phosphoribosyltransferase* (*HPRT*) locus is widely used as a negative selection marker [70, 135]. Inactivation of HPRT by a single round of PGE confers 6-thioguanine resistance in male cells. In our system, a rAAV targeting vector (**Figure 1A**) was assembled to disrupt exon 3 of HPRT (**Figure 1B**) with a neomycin-resistance gene (Neo^R) expression cassette. Following G418 selection, PGE and random integration events could be distinguished based on their 6-thioguanine resistance or sensitivity. In order to differentiate the viral DNA from its chromosomal counterpart, each homology arm of the virus was marked with 4 SNPs that generated unique restriction enzyme recognition sites. In addition, a 22 bp hairpin structure, which is refractory to the MMR machinery [127, 136] that was generated by the inclusion of 3 to 4 SNPs, was also introduced into each homology arm (**Figure 1A**). The homology arms of the targeted and randomly integrated clones could be amplified from the integrated loci (**Figure 1C**) using diagnostic PCRs. Primer pairs P1xP3 and P4xP6 (PGE primers) specifically amplified the left and right homology arms of targeted clones, whereas P2xP3 and P4xP5 (random integration primers) amplified the randomly integrated clones with intact homology arms (**Figure 1C**). The retention of the viral SNPs and hairpins was analyzed either by restriction enzyme sensitivity or DNA sequencing, or both.

PGE is characterized by bidirectional conversion tracts with a linear distribution

To elucidate the molecular mechanism of rAAV-mediated PGE, it was important to characterize which parts of the homology arms were integrated into the genome (the conversion tracts). Since the retention of SNPs can be influenced by MMR, PGE was

initially performed in the MMR-deficient, male HCT116 and DLD-1 cell lines, which are deficient in MLH1/MSH3 and MSH6, respectively [137, 138]. In the later part of this paper we demonstrate that while the MMR status of a cell affects the frequency of PGE it importantly does not affect the SNP retention profile. After rAAV infection, cells were selected with G418 and 6-thioguanine. A total of 230 (for HCT116) and 92 (for DLD-1) correctly targeted clones were confirmed by PCR and analyzed for the retention frequency of viral SNPs, which was then plotted against the position of the SNPs on the homology arms (**Figure 1E, 1F, Tables S1 and S6**). Strikingly, the viral SNPs were retained in a virtually linear gradient pattern: R^2 equaled 0.981 and 0.996 for the left and right homology arms, respectively, in HCT116 cells (**Figure 1E**) and 0.945 and 0.991 for the left and right homology arms, respectively, in DLD-1 cells (**Figure 1F**). The inner SNPs had the highest chance of retention, whereas the outer markers were mostly lost during PGE. The linear distribution of conversion tracts suggested that the positions of HJ resolution were evenly distributed throughout the homology arms because when HJ resolution occurs, the viral homology arms distal to that position will not be retained. Importantly, the linear conversion tracts observed in human PGE contrast with the exponential conversion tracts reported for meiotic recombination in yeast and *Drosophila* and for mitotic recombination in yeast and mouse embryonic stem cells {[123, 131-134]; **Figure S2**}, which implied that the dynamics of HJ formation/resolution during PGE in human somatic cells may be different from similar processes in other organisms.

To determine if the even distribution of HJ resolution was intrinsic to rAAV-mediated PGE or was a general feature of PGE in human cells, a parallel transfection experiment was performed using a dsDNA donor that was identical to rAAV except that it was double-stranded and it did not contain the inverted terminal repeats (**Figure 1D**). Ultimately, 18 correctly targeted clones were recovered despite the extremely low

targeting efficiency of this approach. SNP analysis revealed an indistinguishable linear retention curve (**Figure 1G and Table S2**). Thus, the even distribution of HJ resolution is a general characteristic of PGE in human somatic cells, which led us to believe that rAAV, as a single-stranded virus, may target the human genome in a mechanism similar to dsDNA donors, *i.e.*, via two-end, ends-out DSBR {[64, 122, 126]; **Figure S3**}.

Homology arms remain mostly Intact during rAAV random integration

While PGE is perforce mediated by homology-directed repair, random integration is believed to be mediated by the NHEJ pathways. To test whether PGE and random integration produce different molecular products, 38 random clones were recovered and analyzed. 37 of these clones could be amplified by both sets of random integration primers (**Figure 1C**), indicating that the entire homology arms are almost always retained during random integration. To rule out potential discontinuous homology arm incorporation, a SNP retention analysis was also performed upon the random integration clones. Strikingly, all the SNPs were 100% retained on both arms of the random clones (**Figure 1H and Table S3**), which confirmed that the homology arms were incorporated intact during random integration. This result is consistent with observations that AAV and rAAV viral:chromosomal DNA junctions reside almost exclusively within the viral inverted terminal repeats instead of the homology arms during random integration [139-141]. The retention of intact viral homology arms during random integration, in contrast to the gradient SNP retention that occurred during PGE, unequivocally demonstrated that rAAV-mediated PGE and random integration are mediated by non-overlapping DSB repair pathways.

rAAV-mediated PGE occurs predominantly via DSBR instead of ssDI

While only single-stranded genomes can be packaged into virions, rAAV becomes double-stranded during replication in the host cell [142]. To determine whether viral ssDNA or dsDNA was the major substrate for PGE, a sectoring assay [64, 123, 126] was performed in MMR-deficient HCT116 and DLD-1 cells (**Figure 2A and 2B**). If double-stranded viral substrates are used for PGE via DSBR (**Figure 2A**), both viral strands will be incorporated into a heteroduplex DNA intermediate with unequal length. When this heteroduplex DNA intermediate is resolved by mitosis *in situ*, the two daughter cells will give rise to a heterogeneous colony containing genetically distinct cells that are reciprocally sectored for some of the SNPs on the homology arms (**Figure 2A**). On the other hand, if PGE occurs via ssDI (**Figure 2B**), a single-stranded viral DNA will be annealed into the heteroduplex DNA. Subsequently, the daughter cell lacking the selection marker will be killed during drug selection, whereas the other will grow into a homogenous colony with all the SNPs unsectored (**Figure 2B**). Consequently, the relative contribution of DSBR and ssDI can be expressed as the ratio of the sectored to unsectored colonies produced by rAAV-mediated PGE.

HCT116 and DLD-1 cells were infected and then allowed to grow into colonies *in situ* in G418- and 6-thioguanine-containing medium. An amount of virus was used to make sure that on average only a single colony was formed in each plate. SNP analysis revealed that 74% and 89% of targeted clones in HCT116 and DLD-1, respectively, were sectored on at least one side of the homology arms (**Figure 3, Tables S4 and S6**), consistent with the DSBR model. Considering that this assay is unable to detect short heteroduplex DNA tracts formed between two neighboring SNPs, this result is likely an underestimation of the actual number of sectored colonies. To rule out the possibility

that the sectoring was generated from doublet colonies or two independent ssDI events, 11 clones that were sectored on both arms were subjected to single-cell subcloning. Sequencing analyses demonstrated that 89.6% of the subclones segregated the SNPs with a perfect *trans* configuration (**Table S5**). Since colonies produced by two independent PGE events will have an equal chance to be *trans* or *cis*, the empirically-observed biased *trans:cis* ratio indicated that most colonies were generated by a single DSBR event. Thus, in contrast to popular belief, rAAV-mediated PGE is predominantly mediated by DSBR in human cells. Nevertheless, since a fraction (26% for HCT116 and 11% for DLD-1) of the targeted clones remained unsectored, we cannot rule out the possible involvement of ssDI as a minor pathway.

The efficiency of rAAV-mediated PGE correlates with the DSBR activity

To confirm that rAAV-mediated PGE efficiency correlated with DSBR, and not ssDI, activity, we transfected HCT116 cells with Rad51K133A, a dominant negative form of Rad51 reported to reduce DSBR and concomitantly elevate SSA [134]. Using episomal reporters for either the total non-crossover products from HDR (Figure 4A) or SSA (**Figure 4B**), we confirmed that expression of the dominant negative indeed reduced DSBR and increased SSA in HCT116 cells (**Figure 4C**). Importantly, the rAAV targeting efficiency at the HPRT locus was reduced by 6.2-fold upon Rad51K133A transfection, which correlated well with the reduced DSBR activity and not the increased SSA activity in these cells (**Figure 4C**). Thus, consistent with the sectoring assay, this result further confirmed that rAAV-mediated PGE is mediated predominantly by DSBR instead of ssDI in human cells.

The process of rAAV-mediated PGE is inverted in the presence of DSBs

Spontaneous endogenous DSBs occur around 10 times per mammalian cell per day [15]. The likelihood that one of these DSBs must be introduced near a target locus in order for rAAV-mediated PGE to occur is statistically improbable. rAAV-mediated PGE must, therefore, employ a mechanism that is independent of the formation of chromosomal DSBs (**Figure 2A**). Nevertheless, rAAV-mediated PGE can be stimulated dramatically by the presence of chromosomal DSBs near the target locus [51-53]. The mechanistic basis for this increase is, however, not understood. To investigate this issue, rAAV was used to "knock-in" an *I-SceI* enzyme recognition sequence onto the X chromosome at a site that corresponded to a position (nt 266), just to the right of the *SacI* (nt 261) site, on the right homology arm of the HPRT rAAV targeting vector (**Figure 5A, 5B and Figure S4**). After transfection with an *I-SceI* expression plasmid, chromosomal DSBs were quantified by ligation-mediated PCR {[143]; **Figure 5D**}. DSBs were detectable 16 hr after transfection, and peaked ~24 hr after transfection (**Figure 5E**). Accordingly, rAAV infections were performed either 12 or 20 hr after *I-SceI* transfection in an attempt to coordinate the viral infection with the chromosomal DSB induction. The absolute PGE efficiency increased by 477- and 582-fold, respectively, in the presence of *I-SceI* (**Figure 5F**), which was consistent with previous reports [51-53]. The random integration frequency was virtually unperturbed by the expression of *I-SceI* (**Figure 5F**). The retention of viral SNPs was then analyzed in 64 targeted clones. Strikingly, the *SspI* and *SacI* sites on the right homology arm were both retained at 100% frequency (**Figure 5G and Table S7**), which was in stark contrast to the linear gradient of SNP loss in spontaneous PGE (compare **Figure 5G** with **Figure 1E and 1F**). The SNPs to the right of the *I-SceI* site (the RHP, *XbaI* and *SbfI*) were lost in a sharper, but

nonetheless linear, gradient (**Figure 5G**). To confirm this finding, we constructed another cell line in which rAAV was used to knock-in an *I-SceI* enzyme recognition sequence into the X chromosome at a site that corresponded to a position (nt -569), just to the left of the *NcoI* (nt -547) site, on the left homology arm of the HPRT rAAV targeting vector (**Figure 5A, 5C and Figure S4**). The rAAV-mediated PGE frequency was also elevated by concomitant *I-SceI* expression (**Figure 5F**). The retention of viral SNPs was then analyzed in 48 targeted clones. In a strikingly mirrored fashion, the *AseI* and *NcoI* sites on the left homology arm were both retained at 100% frequency, while the SNPs to the left of this region (the LHP, *EcoRI*, *NdeI*) were lost in a linear gradient (**Figure 5H and Table S8**).

The plateaued conversion tracts observed in these 2 experiments are predicted from an "ends-in" DSBR model in which recombination is initiated not by the vector DNA but by the broken chromosome (**Figure 6A**). In contrast to spontaneous rAAV PGE where the viral DNA "attacks" the unbroken chromosome in an ends-out configuration (**Figure 2A**), in DSB-induced PGE the broken chromosomal ends are instead processed and invade the virus in an ends-in configuration (**Figure 6A and Figure S3**). Without drug selection, the random distribution of HJ resolution would produce linear conversion tracts peaking at the *I-SceI* site (**Figure 6A-1**; cartooned for the rightward *I-SceI* site). However, because G418 selection was imposed, any HJs that were resolved between the *I-SceI* site and the selection cassette would have been lost. Consequently, the initiation of recombination with the chromosomal *I-SceI*-restricted ends and the requirement for the retention of the viral selection cassette precisely explain the conversion tracts that we obtained (compare **Figure 5G** with **Figure 6A-2**). In summary, the introduction of a chromosomal DSB inverts the process of PGE such that the viral DNA becomes the "attackee" instead of the attacker.

These data also established an important corollary. Three pathways can act independently to resolve an HDR intermediate: the DSBR model, HJ dissolution and SDSA {[144, 145]; **Figure 6**}. DSBR features the formation and resolution of double HJs (**Figure 6A**) whereas inward branch migration of the HJs can cause HJ dissolution (**Figure 6B**). Alternatively, in SDSA the synapse collapses before the formation of the second HJ (**Figure 6C**). SDSA is believed to be the major pathway of mitotic recombination in yeast and plants [146, 147]. It is also the preferred pathway of repairing an *I-SceI*-induced DSB in mouse and human cells [33]. Importantly, both the SDSA and HJ dissolution models predict unidirectional conversion tracts with the retention of one half of the *I-SceI* site and the loss of all of the SNPs rightward of the right *I-SceI* site (**Figure 6-3**), or leftward of the left *I-SceI* site (not shown), a minor pattern that was observed in only 17% of the clones (**Table S7**). Collectively, these results suggest that although SDSA may be the major pathway for recombination in mitotic cells, the DSBR model is the predominant form of HDR that leads to PGE in human somatic cells using rAAV or dsDNA donors with central heterology.

The strong inhibitory role of MLH1 on PGE via its anti-recombination activity

Since the SNPs engineered into the rAAV targeting vector generated mismatches in the heteroduplex DNA intermediate, we wished to assess if they were sensitive to MMR. Thus, another rAAV targeting vector was constructed with only 2 SNPs and tested in the parental HCT116 (MMR-deficient) cell line (Figure 7A). The targeting efficiency was 7.5-fold higher compared to the original vector, which contained 15 SNPs (Figure 7B). These data indicated that the presence of mismatches deleteriously affected PGE even in a MMR-reduced background, a result that can be attributed to the residual MMR activity present in this cell line [148]. To further address the role of the

MMR system, PGE was performed in an MMR-proficient variant ($MLH1^+$), in which the mutated MLH1 gene in HCT116 cells was corrected by rAAV-mediated knock-in (Figure 7B, inset). Targeting efficiency decreased by more than 50-fold in $MLH1^+$ cells for each of the vectors respectively compared with the isogenic MLH1-defective parental line (Figure 7B). Collectively, these data demonstrated that the MMR gene MLH1 exerts a strong inhibitory effect on PGE [123, 149].

Mismatch repair has two well-documented activities. One is as a "spell-checker" to correct post-replication mismatches in DNA and the other is as an "anti-recombinase", by impeding the formation of homeologous heteroduplex DNA [149, 150]. To assess which of these two activities was responsible for reducing PGE, 20 targeted clones were recovered — despite the extremely low targeting efficiency in $MLH1^+$ cells — and analyzed for SNP retention (Figure 7C and Table S9). Importantly, the SNP retention curve for $MLH1^+$ cells was indistinguishable from the parental ($MLH1^-$) linear retention curve (compare Figure 1E and F with Figure 7C). Moreover, the hairpins, which are refractory to the spell-checking activity of MMR [127, 148], were retained at the same frequency as is predicted by the linear regression of other SNPs, which are substrates for spell-checking. Finally, the percentage of discontinuous gene conversion tracts (a hallmark of spell-checking) did not change significantly in the MMR-proficient, compared to the MMR-deficient, background (compare Table S9 with Table S1, respectively). These results demonstrated that the presence of MLH1 exercised no detectable spell-checker activity upon the mismatches in the heteroduplex DNA intermediate and implied that the large, negative impact of MLH1 on PGE was instead due to anti-recombination activity of MMR [123, 149, 150]. Finally, to test whether the MMR system affects random integration, 22 G418-resistant 6-thioguanine-sensitive clones were recovered from the $MLH1^+$ background and analyzed for SNP retention. All but one of them could

be amplified using the random integration primers, and once again, 100% of the viral SNPs were retained (Figure 7D and Table S10), which is consistent with the observation that MMR does not affect NHEJ [150].

Discussion

rAAV uses the DSBR model of HDR for PGE

Although rAAV is widely used in laboratory and clinical studies, the mechanism of rAAV-mediated PGE has remained obscure. Since rAAV is packaged exclusively as a single-stranded virus, several reports have suggested that rAAV-mediated PGE is mediated by ssDI [42, 43]. Moreover, the ssDI model is supported by indirect evidence that minute virus of mouse, a related parvovirus, shows a strand-specific bias in PGE [151]. Our data, however, using three lines of evidence demonstrate that rAAV-mediated PGE is mediated by the DSBR model of HDR using double-stranded viral substrates: (1) rAAV-mediated PGE produces conversion tracts indistinguishable from those of dsDNA-mediated PGE, which is dictated by two-end, ends-out DSBR [64, 126]. (2) rAAV-mediated PGE is associated with the formation of sectored colonies in a *trans* configuration, which is characteristic of the DSBR model. (3) rAAV-mediated PGE frequency correlated with DSBR, and not SSA, activity through the use of Rad51K133A transfections. These results demonstrate that rAAV has to become double-stranded — either by host DNA polymerases or by annealing of the plus and minus viral strands — before targeted integration can occur.

What is less clear, given that rAAV uses the same mechanism as dsDNA donors, is why rAAV targets human cells so much more robustly. We suggest that there are several viral elements of rAAV that may positively influence PGE. For example, the capsid proteins may facilitate virus transduction and nuclear trafficking via interaction with cellular receptors [152] to generate higher nuclear concentrations of the viral DNA versus transfected DNA. In addition, the hairpin-structured inverted terminal repeats may serve as physical barriers to protect the ends of the viral genome from nuclease

degradation during nuclear trafficking. An alternative possibility that we favor is that the inverted terminal repeats may facilitate the formation of active recombination substrates. Thus, besides the recombinogenic linear viral dsDNA, infected cells also contain a mixture of viral ssDNA along with circular and concatemerized dsDNA [153]. Our ends-out recombination model requires that both ends of the viral genome are accessible to exonuclease resection, which means that the linear, monomeric double-stranded viral genomes are the only active substrates that can be used for PGE. Since the inverted terminal repeats suppress the intra- and intermolecular recombination that generates viral circular and concatemerized dsDNA [154], they may facilitate PGE by favoring the existence of the active recombination substrates. On the contrary, dsDNA-mediated PGE vectors may be efficiently inactivated by circularization or concatemerization before PGE can occur. Needless, to say, none of these hypotheses are mutually exclusive and they may act synergistically to enhance rAAV-mediated PGE.

rAAV-mediated PGE as a model to study DSBR in human somatic cells

The locations of crossovers are determined by the initial positions of HJ formation and branch migration activity. Comprehensive gene conversion analyses have been performed in yeast, flies and mouse embryonic stem cells, which revealed short conversion tracts with exponential distribution during meiotic and mitotic HDR {[123, 131-133]; **Figure S2**}. These studies indicated that the crossovers were more likely to be resolved or dissolved near the initiation site of strand invasion, probably as a result of branch migration. Although similar studies have been undertaken in mammalian systems [33, 125-128, 155] the generality of the conclusions were restricted by the limited scale of the data. Taking advantage of the high targeting efficiency of rAAV, we performed a conversion tracts analysis for spontaneous (**Figure 1E and 1F**) and DSB-

induced (**Figure 5G and 5H**) PGE in human cells with unprecedented resolution. In contrast to previous studies, we obtained long conversion tracts with linear distribution, indicating that crossovers are evenly, and not exponentially, distributed along the homology arms in the DSBR model. We further confirmed the generality of the conversion tracts using dsDNA-mediated PGE, although on a smaller scale (**Figure 1G**). Assuming that each segment of the homology arms has the same tendency to initiate strand exchange [156], we propose that the linear conversion tracts in human cells are shaped primarily by the even distribution of HJ formation and is minimally impacted by branch migration in the DSBR model. It should be noted that alternative scenarios are possible. For example, rather than formation of a second HJ (**Figure 2A**), the distal ends could be resolved by cleavage with structure-specific endonucleases such as XPF/ERCC1 [157, 158]. Our linear conversion tracts favor the former scenario, but we cannot rule out the latter possibility.

Branch migration reshapes the distribution of crossovers and determines the amount of genetic information exchanged during HDR. Interestingly, bacterial RecA and its mammalian Rad51 homologs facilitate branch migration in different directions: RecA moves the HJs away from DSBs to encourage the exchange of genetic material in bacteria, whereas in lower eukaryotes, Rad51 shifts the HJs towards DSBs to minimize gene conversion tracts [159]. Our results are consistent with the *in vitro* observation that the branch migration activity of human Rad51 is substantially lower than its yeast counterpart [159], which suggests that human cells may have adopted an energy-saving strategy to repair somatic DSBs by HDR without suppressing the amount of genetic material in the exchange.

An additional insight from our studies is the demonstration that meganucleases stimulate PGE by promoting chromosome-initiated, ends-in DSBR. Creating a DSB in a

target locus increases the frequency of PGE by 2 to 3 orders of magnitude {[51-53, 160]; **Figure 5F**}, which makes artificial meganucleases promising tools for genetic engineering. The mechanism for the enhanced PGE frequency was, however, unknown. Importantly, our chromosome-initiated ends-in DSBR model immediately provides an explanation for this profound enhancement. As discussed earlier, the viral DNA inside an infected cell can exist as linear, circular or concatemeric species and only the former of these is proficient for ends-out recombination. Since the majority of the viral genomes are converted into circular or concatemeric forms by cellular DSB repair pathways shortly after infection [142, 153] the efficiency for spontaneous PGE is low. In contrast, in DSB-induced PGE, the broken chromosome ends can invade all of these exogenous species to initiate HDR. Also, this ends-in DSBR involves the resolution of only two — instead of the four — HJs required for the ends-out DSBR. These differences may together contribute to the orders of magnitude increase in targeting efficiency.

The demonstration of two modes for PGE explains an additional conundrum in the field. Thus, by themselves, ODNs are poor donors for PGE in mammalian cells [161]. Paradoxically, with the development of artificial meganucleases, ZFNs [85], TALENs [86] and CRISPR/Cas9 [162], there has been a spate of recent papers demonstrating that single-stranded oligonucleotides can be efficiently used to facilitate HDR in the presence of a DSB {*e.g.*, [163, 164]}. This “paradox” however, is precisely what our data would predict: by itself, ODNs would need to engage an endogenous genomic lesion to initiate PGE via the minor HDR pathways (*e.g.*, ssDI). In contrast, following a meganuclease-induced DSB, the resulting chromosomal ends should efficiently and productively be able to interact with accompanying ODN donors via SDSA.

Finally, our data demonstrate that the DSBR model is the preferred HDR pathway leading to PGE using rAAV or dsDNA donors with central heterology in human cells.

The DSBR model has become the paradigm of HDR in yeast [130], which is characterized by the formation of double HJs and resolution by resolvases (**Figure 6A**). However, this model has been challenged in mammalian cells by the fact that mitotic recombination is infrequently associated with crossovers. SDSA emerged as an alternative model [165], in which the invading strand anneals back with its original partner after *de novo* DNA synthesis without the formation of HJs (**Figure 6C**). In plants and mammals, a large body of evidence suggests that SDSA is the preferred pathway of mitotic recombination [33, 146, 147]. There is less convincing data that SDSA is utilized for PGE, although it should be noted that the ERCC1/XPF nuclease complex, which has documented roles in SSA and in SDSA [158], can impact the process of mammalian PGE as well [166]. Our conversion tracts analysis in the presence of a chromosomal DSB, however, indicated that the bulk of the PGE products are generated by the DSBR model, at least when the conversion of a large drug selection marker is required (**Figure 6A**). This result strongly argues that — in human somatic cells — PGE using dsDNA donors, including rAAV, is most accurately described by the DSBR model.

In toto, it should also be emphasized that there are a multitude of differences, some subtle and some not so that distinguish human PGE from that described in other systems: *e.g*., 1) the even distribution of crossovers in the homology arms, 2) the preferred use of DSBR versus SDSA or HJ dissolution, and 3) the preferential use of broken chromosomal ends over the ends of exogenous DNA. Understanding the mechanistic underpinnings of these differences will be critical to improve the efficacy of PGE for therapeutic purposes.

Figures and Legends

Figure 1. PGE is characterized by linear conversion tracts

(A) The rAAV PGE donor. The Neo^R cassette (white rectangle) is flanked by the homology arms (green and blue rectangles). NdeI, EcoRI, NcoI, AseI, SspI, SacI, XbaI and SbfI represent vector-borne restriction sites created by SNPs. LHP/RHP represent vector-borne palindromes (lollipops) created by introducing SNPs. The flanking hairpins represent inverted terminal repeats. (B and C) The HPRT locus before and after PGE. The Neo^R cassette replaces exon 3 (grey) of HPRT upon PGE. The theoretical positions of the viral markers are indicated in bold vertical lines and (?) symbols. The arrows represent PCR primers. P1xP3 and P4xP6 amplify the left and right homology arms of the gene targeted clones, respectively, and P2xP3 and P4xP5 amplify the homology arms of the randomly integrated clones, respectively. The LHP destroys a chromosomal BbvCI restriction site upon integration. (D) The dsDNA PGE donor. All symbols are defined above. (E, F, G and H) SNP retention signatures of rAAV-mediated PGE for HCT116 and DLD-1 cell lines, dsDNA-mediated PGE and random integration, respectively. The distance to the central heterology (cartooned as a vertical black line) is calculated from the inner ends of the homology arms. Markers on the left homology arms are indicated with negative distances. Green and purple lines represent the linear regression between the retention frequency and the distance of the viral markers for the left and right homology arms, respectively.

Figure 1

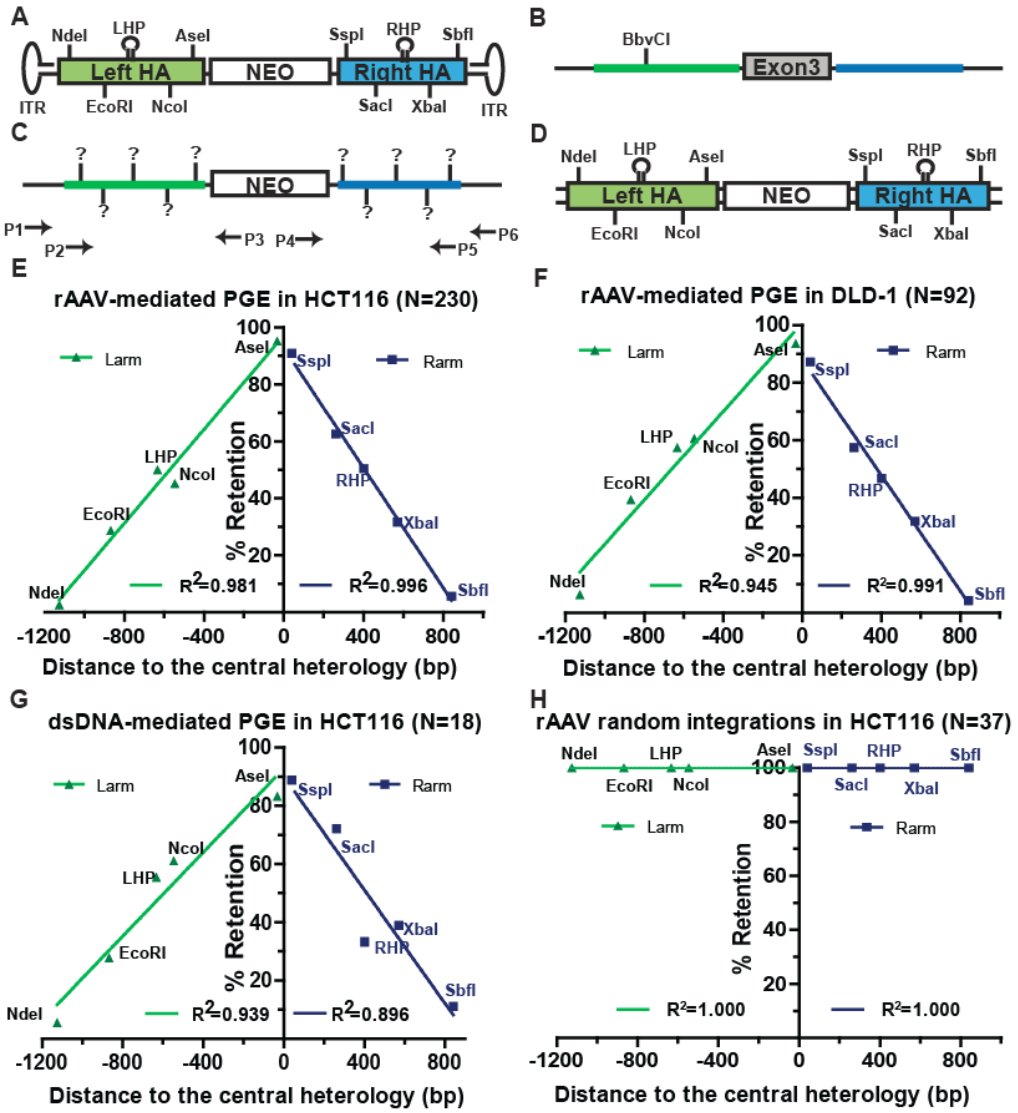

A
NdeI
LHP
AseI
SspI
RHP
SbfI
Left HA
NEO
Right HA
ITR
EcoRI
NcoI
SacI
XbaI
ITR
B
BbvCI
Exon3
C
NEO
P1
P2
P3
P4
P5
P6
D
NdeI
LHP
AseI
SspI
RHP
SbfI
Left HA
NEO
Right HA
EcoRI
NcoI
SacI
XbaI
E
rAAV-mediated PGE in HCT116 (N=230)
Larm
Rarm
% Retention
Distance to the central heterology (bp)
R2=0.981
R2=0.996
F
rAAV-mediated PGE in DLD-1 (N=92)
R2=0.945
R2=0.991
G
dsDNA-mediated PGE in HCT116 (N=18)
R2=0.939
R2=0.896
H
rAAV random integrations in HCT116 (N=37)
R2=1.000
R2=1.000

Figure 2. Models for rAAV-mediated PGE

(A) The DSBR model. Black, red and blue lines correspond to genomic DNA, viral and genomic homology arms, respectively; the bold green line corresponds to the selection cassette. The vertical arrows imply that the viral DNA becomes double-stranded and the inverted terminal repeats are processed before integrating into the genome. Open arrows represent the sites of HJ cleavage and ligation. A sectored colony is formed during mitosis. (B) The ssDI model. All symbols are as in (A). The virus that anneals to the genomic DNA is single-stranded. In the ensuing mitosis, an unsectored colony is formed under drug selection.

Figure 2

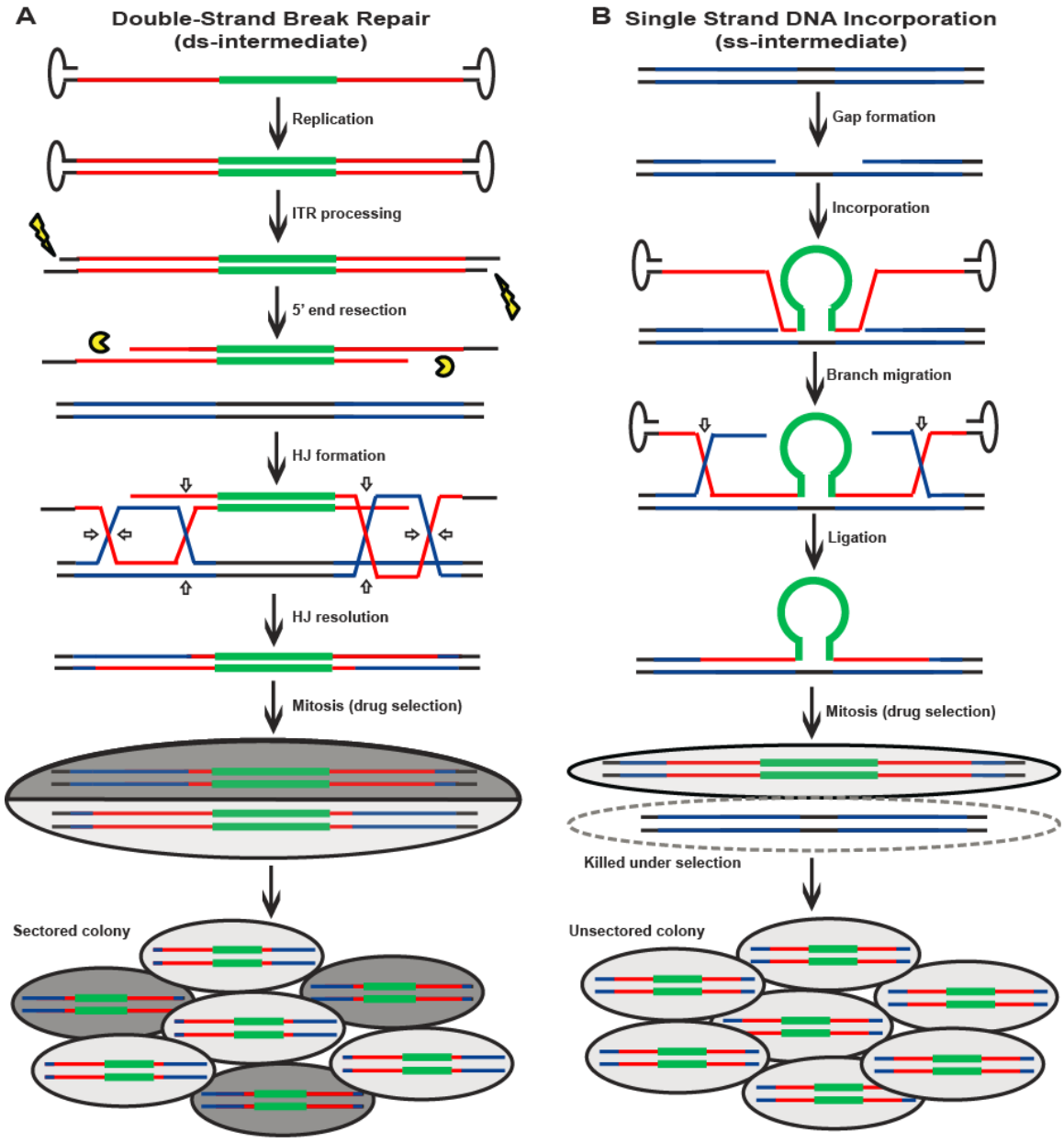

A
Double-Strand Break Repair
(ds-intermediate)
Replication
ITR processing
5' end resection
HJ formation
HJ resolution
Mitosis (drug selection)
Sectored colony
B
Single Strand DNA Incorporation
(ss-intermediate)
Gap formation
Incorporation
Branch migration
Ligation
Mitosis (drug selection)
Killed under selection
Unsectored colony

Figure 3. rAAV-mediated PGE forms sectored colonies

Solid boxes on the top (not to scale) represents diagnostic markers on the virus (blue) and genomic DNA (yellow). The numbers indicate the actual positions of the markers. The Neo^R cassette and exon 3 of HPRT are indicated in white and grey, respectively. Each line on the bottom corresponds to an independent PGE event. The blue, yellow and green segments are color-coded to represent viral, genomic and sectored tracts, respectively. The top and bottom panels show results obtained from HCT116 and DLD-1 cells, respectively.

Figure 3

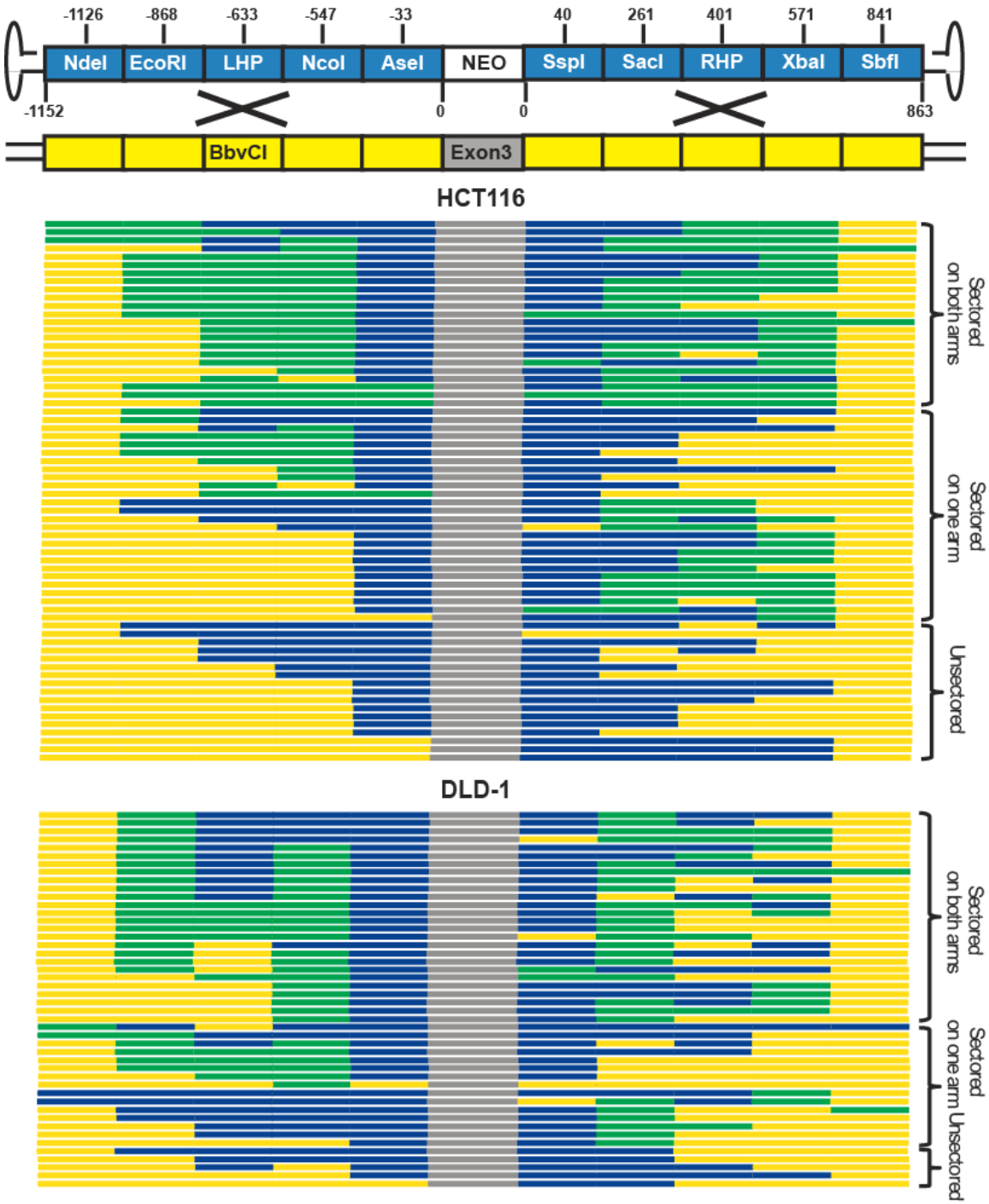

-1126
-868
-633
-547
-33
40
261
401
571
841
NdeI
EcoRI
LHP
NcoI
AseI
NEO
SspI
SacI
RHP
XbaI
SbfI
-1152
0
0
863
BbvCI
Exon3
HCT116
Sectored on both arms
Sectored on one arm
Unsectored
DLD-1
Sectored on both arms
Sectored on one arm
Unsectored

Figure 4. The efficiency of rAAV PGE correlates with the DSBR activity

(A) The DSBR assay. SceGFP is a full-length GFP gene disrupted by an I-SceI site. DSBR between SceGFP and the internal GFP (iGFP) fragment on the same plasmid upon I-SceI digestion restores GFP activity. (B) The SSA assay. 5'GFP and 3'SceGFP are GFP fragments bearing 266 bp of homology. SSA repair of the I-SceI-induced DSB generates a functional GFP gene. (C) The efficiency of DSBR, SSA and rAAV-mediated PGE. The indicated cell lines were analyzed using the DSBR and SSA assays as well as for rAAV-mediated PGE. The mean ± SEM of three independent experiments is shown.

Figure 4

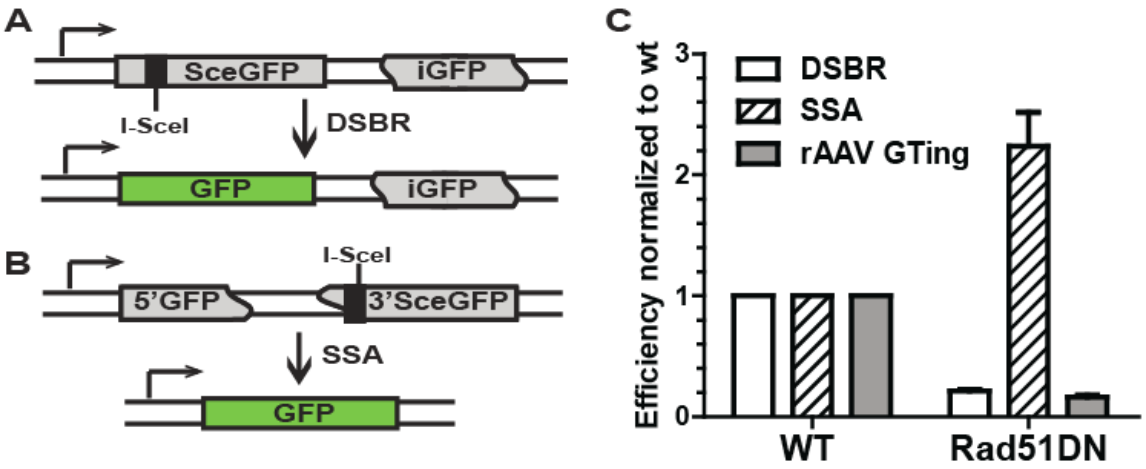

A
SceGFP
iGFP
I-SceI
DSBR
GFP
iGFP
B
I-SceI
5'GFP
3'SceGFP
SSA
GFP
C
Efficiency normalized to wt
0
1
2
3
DSBR
SSA
rAAV GTing
WT
Rad51DN

Figure 5. DSB-induced PGE generates plateaued conversion tracts

(A) The rAAV PGE donor. See the legend to Figure 1A for details. (B) The HPRT locus. An I-SceI restriction site (orange) was knocked-in to the chromosome at the indicated corresponding vector position. The lightning bolt denotes that DSBs can be induced upon I-SceI expression. (C) The HPRT locus; left side analysis. All symbols as in (B). (D) Scheme for the LM PCR. A linker with a 5' I-SceI overhang was ligated to the 3' end of the genomic I-SceI-generated break. The presence of a ligation product was quantitated with the primers indicated by arrows. (E) I-SceI-induced chromosomal DSBs can be detected within 24 hr of I-SceI expression. A gel electrophoresis analysis of the PCR products generated using genomic DNA isolated at the indicated times following I-SceI expression. β-Actin was used as a loading control. (F) The efficiency of I-SceI-induced rAAV PGE. Cells were infected with rAAV without (Ctrl) or 12 or 20 hr (for the right side) or 24 hr (for the left side) after I-SceI expression. The PGE and random integration frequencies were normalized to the no I-SceI control. (G) The conversion tracts of I-SceI-induced rAAV-mediated PGE; right side. The dotted orange line indicates the position of the I-SceI site. All other symbols are defined in Figure1A. (H) The conversion tracts of I-SceI-induced rAAV-mediated PGE; left side. All symbols are as in (G).

Figure 5

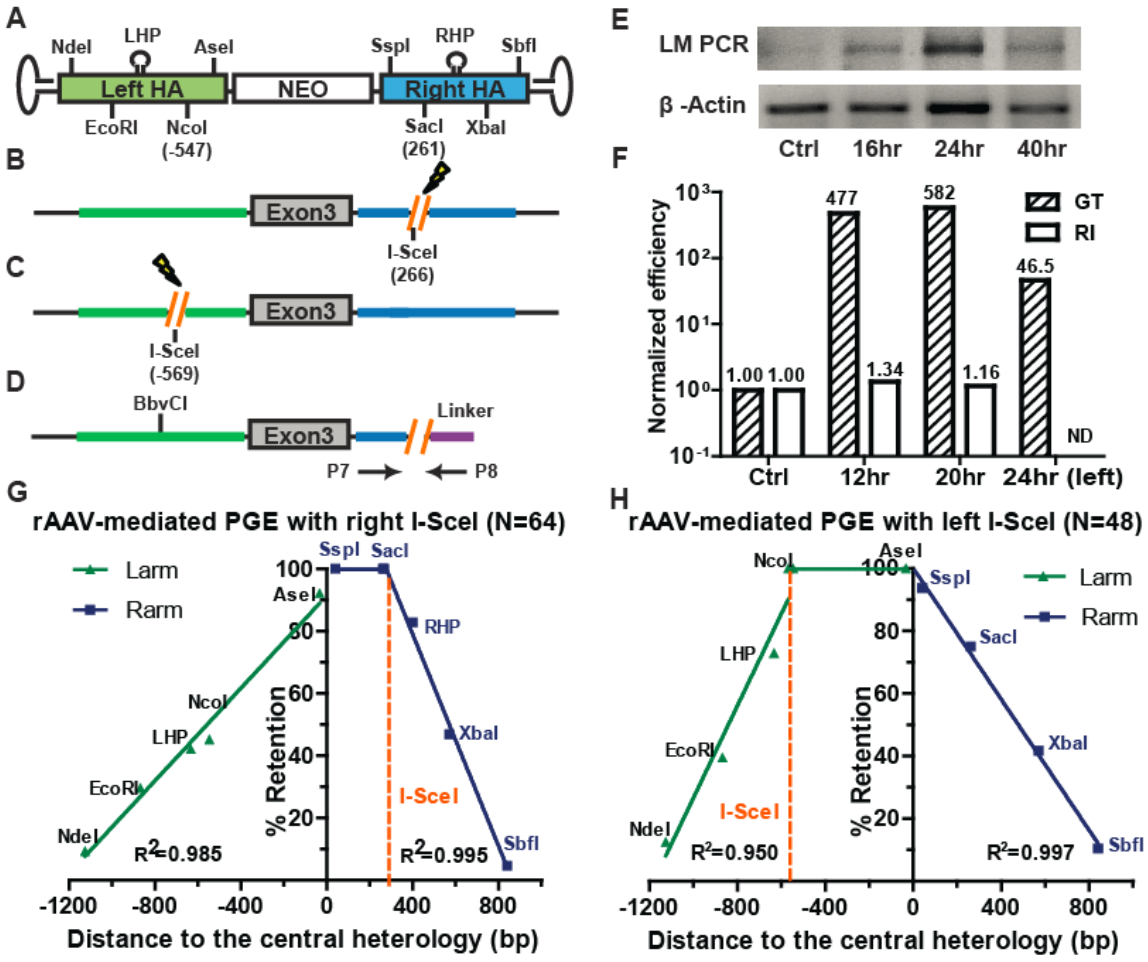

A
NdeI
LHP
AseI
SspI
RHP
SbfI
Left HA
NEO
Right HA
EcoRI
NcoI
(-547)
SacI
(261)
XbaI
B
Exon3
I-SceI
(266)
C
Exon3
I-SceI
(-569)
D
BbvCI
Exon3
Linker
P7
P8
E
LM PCR
β -Actin
Ctrl
16hr
24hr
40hr
F
Normalized efficiency
GT
RI
1.00
1.00
477
1.34
582
1.16
46.5
ND
Ctrl
12hr
20hr
24hr (left)
G
rAAV-mediated PGE with right I-SceI (N=64)
Larm
Rarm
% Retention
I-SceI
R2=0.985
R2=0.995
Distance to the central heterology (bp)
H
rAAV-mediated PGE with left I-SceI (N=48)
R2=0.950
R2=0.997
Distance to the central heterology (bp)

Figure 6. Models for DSB-induced rAAV PGE

(A) rAAV-mediated PGE in the presence of DSBs. Dotted lines and arrowheads correspond to de novo DNA synthesis, which is color-coded to match the templates. Orange slashes represent half I-SceI sites and the lightning bolt represents I-SceI-induced cleavage. All other symbols are as in Figure 2. The chromosomal ends are processed and invade the viral DNA in an ends-in configuration. Two predicted SNP retention patterns (minus and plus drug selection, respectively) are cartooned as (1) and (2), respectively. (B) HJ dissolution. Branch migration forces the HJs towards the drug selection cassette and the HJ is cancelled. The predicted SNP retention pattern is cartooned in (3). (C) SDSA. If the synapsed structure shown in (A) collapses, recombination can still occur by SDSA. This mechanism, like HJ dissolution (B), predicts the SNP retention pattern shown in (3).

Figure 6

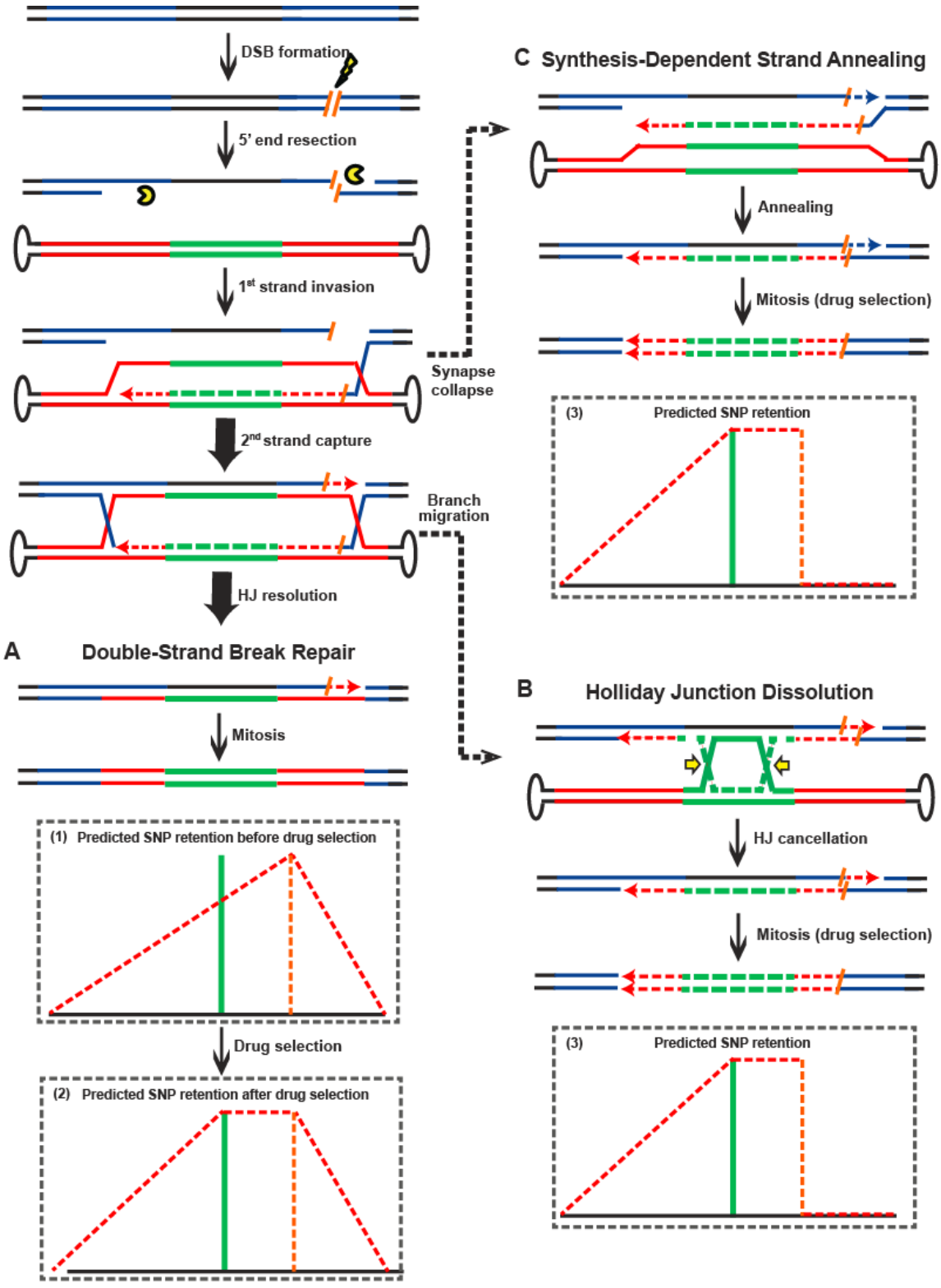

DSB formation
5' end resection
1st strand invasion
Synapse collapse
2nd strand capture
Branch migration
HJ resolution
A Double-Strand Break Repair
Mitosis
(1) Predicted SNP retention before drug selection
Drug selection
(2) Predicted SNP retention after drug selection
C Synthesis-Dependent Strand Annealing
Annealing
Mitosis (drug selection)
(3) Predicted SNP retention
B Holliday Junction Dissolution
HJ cancellation
Mitosis (drug selection)
(3) Predicted SNP retention

Figure 7. rAAV PGE is suppressed in a mismatch repair-proficient background

(A) The rAAV PGE donors. All symbols as in Figure1A. 2SNPs and 15SNPs indicate the total number of mismatches in the vectors. (B) Effects of mismatches and the host mismatch repair status on rAAV mediated PGE. The efficiency of rAAV mediated PGE is expressed as the ratio of PGE events divided by the sum of the PGE plus randomly integrants. All results are normalized to the parental (MLH1-) cell line. The mean ± SEM of three independent experiments is shown. The MLH1 expression in the parental and MLH1+ cell lines is shown in the inserted Western blot panel. b-actin was used as a loading control. (C and D) The SNP retention signature of rAAV mediated PGE and random integration, respectively, in the mismatch repair (MLH1+) - proficient background. All symbols are as in **Figure 1E**.

Figure 7

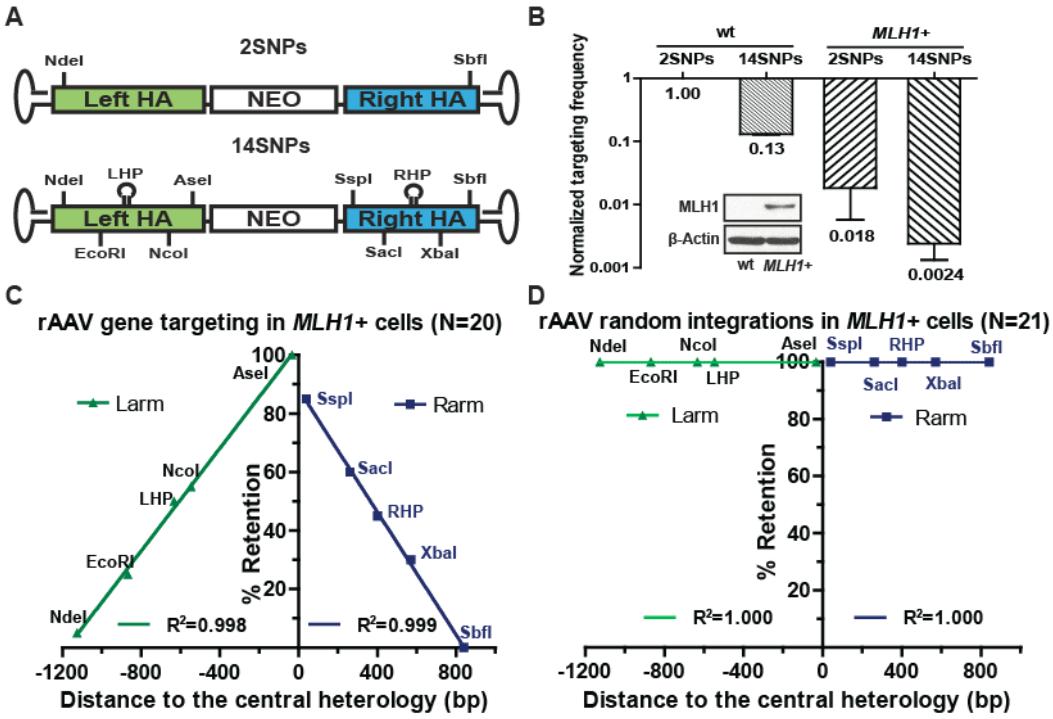

A
2SNPs
NdeI
SbfI
Left HA
NEO
Right HA
14SNPs
NdeI
LHP
AseI
SspI
RHP
SbfI
Left HA
NEO
Right HA
EcoRI
NcoI
SacI
XbaI
B
wt
MLH1+
2SNPs
14SNPs
2SNPs
14SNPs
Normalized targeting frequency
1
0.1
0.01
0.001
1.00
0.13
0.018
0.0024
MLH1
β-Actin
wt
MLH1+
C
rAAV gene targeting in MLH1+ cells (N=20)
Larm
Rarm
% Retention
Distance to the central heterology (bp)
AseI
NcoI
LHP
EcoRI
NdeI
SspI
SacI
RHP
XbaI
SbfI
R²=0.998
R²=0.999
D
rAAV random integrations in MLH1+ cells (N=21)
NdeI
NcoI
AseI
SspI
RHP
SbfI
EcoRI
LHP
SacI
XbaI
Larm
Rarm
% Retention
R²=1.000
R²=1.000
Distance to the central heterology (bp)

Figure S1. Overview of rAAV production and PGE

A cartoon strategy for rAAV virus production and PGE is shown. At the top left are cartooned three plasmids that contain *i*) the AAV viral genes: Rep (orange rectangle) for replication and Cap (purple rectangle) for capsid, *ii*) the rAAV vector containing a backbone (purple lines), the ITRs (bubbles), HAs (red lines) and the selection cassette (green box) and *iii*) the plasmid encoding adenoviral (Ad) helper functions (gray rectangle). These three plasmids are triple transfected into AAV293 cells, where the viral and Ad helper proteins are expressed and facilitate the replication of the viral DNA. Virions (hexagons), containing single-stranded rAAV, are subsequently collected from the supernatant of these cells and used to infect a target cell. In the target cell, the rAAV once again assumes a double-stranded form and facilitates PGE at your favorite gene (YFG).

Figure S1

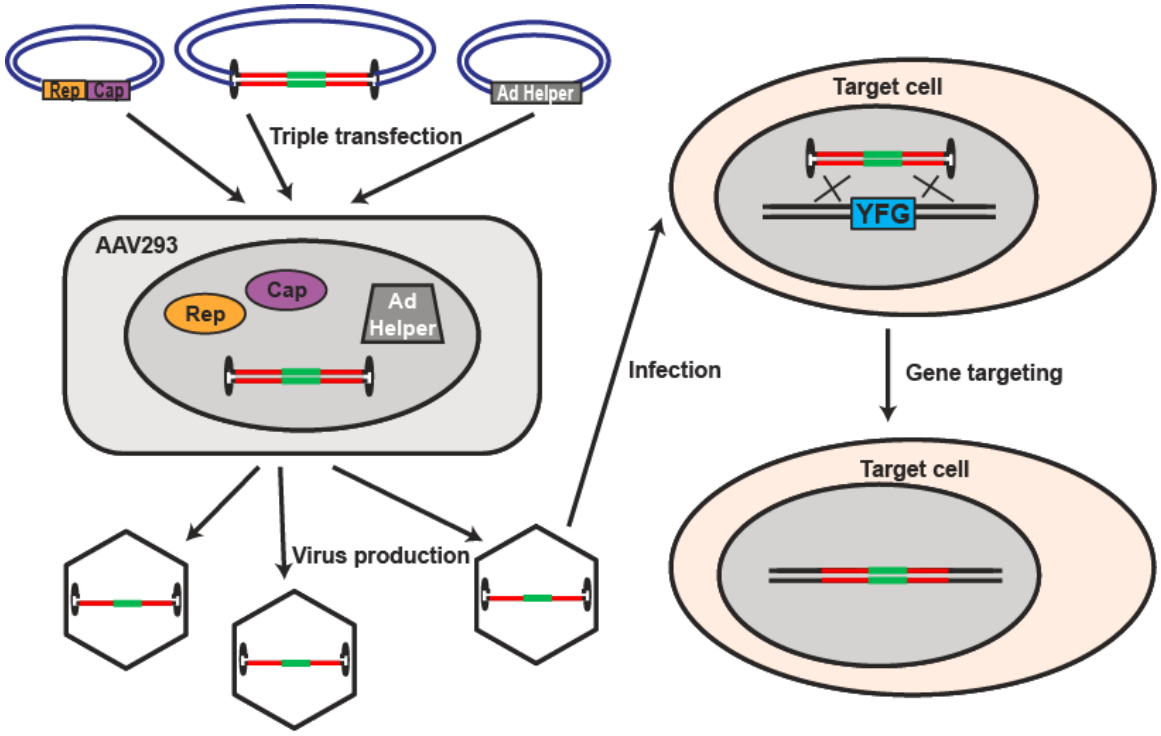

Rep
Cap
Ad Helper
Triple transfection
AAV293
Cap
Rep
Ad Helper
Virus production
Infection
Target cell
YFG
Gene targeting
Target cell

Figure S2. Schematic illustrations of the difference in conversion tracts of lower organisms and human somatic cells

In all organisms, the process of PGE appears to be initiated with the same steps: strand resection (PacMan™) and HJ formation. The blue lines represent chromosomal DNA, the red lines viral DNA and the green lines the selection cassette. In yeast, flies and murine ES cells, the process of branch migration (orange arrows) pushes the HJs towards each other (left), whereas in human cells this process is apparently negligible. HJ resolution (small white arrows) of these structures generates either an exponential SNP retention curve (in the presence of inward branch migration, left) or a linear SNP retention curve (in the absence of branch migration, right). The amount of genetic information that is exchanged is correspondingly restricted (left) or enlarged (right).

Figure S2

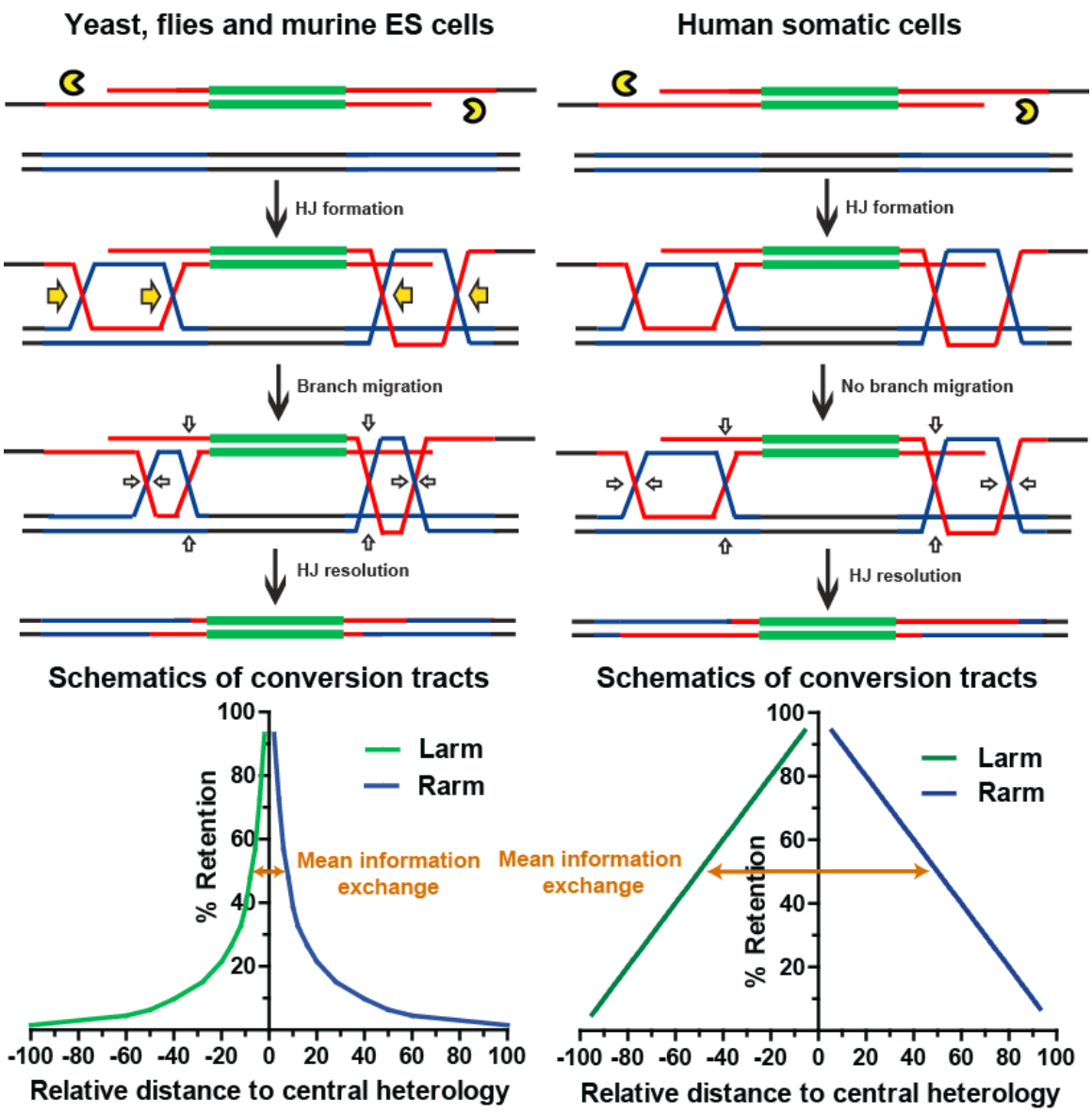

Yeast, flies and murine ES cells
Human somatic cells
HJ formation
HJ formation
Branch migration
No branch migration
HJ resolution
HJ resolution
Schematics of conversion tracts
Schematics of conversion tracts
Larm
Rarm
% Retention
Mean information exchange
Mean information exchange
Relative distance to central heterology
Relative distance to central heterology

Figure S3. Definitions of ends-out and ends-in recombination

A cartoon of a chromosome (blue oval with yellow circular centromere) and a PGE vector (red lines) containing a drug selection cassette (green rectangle) is shown. In the ends-out recombination, the 3' ends of the targeting vector invade (black lines with arrowheads) — in directions opposite to each other — the chromosome in separate HDR reactions. In the spontaneous ends-in recombination, the 3' ends of the targeting vector invade the chromosome in directions facing each other. In the chromosomal DSB-induced ends-in recombination, the broken chromosomal ends (jagged blue ovals) invade/anneal to the targeting vector in directions facing each other.

Figure S3

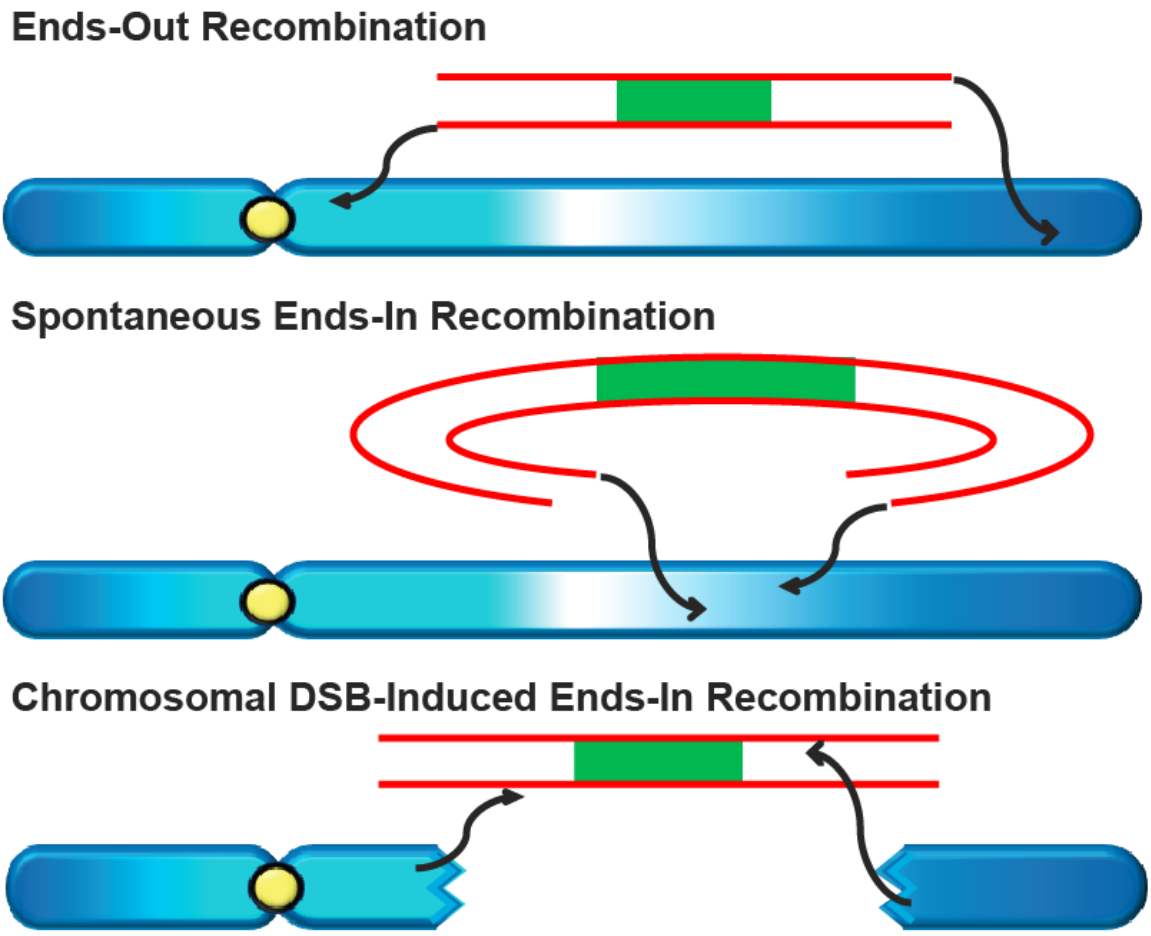

Ends-Out Recombination
Spontaneous Ends-In Recombination
Chromosomal DSB-Induced Ends-In Recombination

Figure S4. Knock-in of I-SceI sites into the HPRT locus

A cartoon of a rAAV *I-SceI* knock-in PGE donor is shown on the top line. The *I-SceI* recognition site is shown a double-hatched line. A cartoon of the relevant portion of the HPRT locus (horizontal line with a rectangular exon 3) is shown on the line below. Following correct GTing, both the *I-SceI* recognition site and the Neo^R drug resistance gene are integrated at the HPRT locus. Following Cre recombination, the Neo^R gene is removed and a solo LoxP scar (shaded triangle) remains.

Figure S4

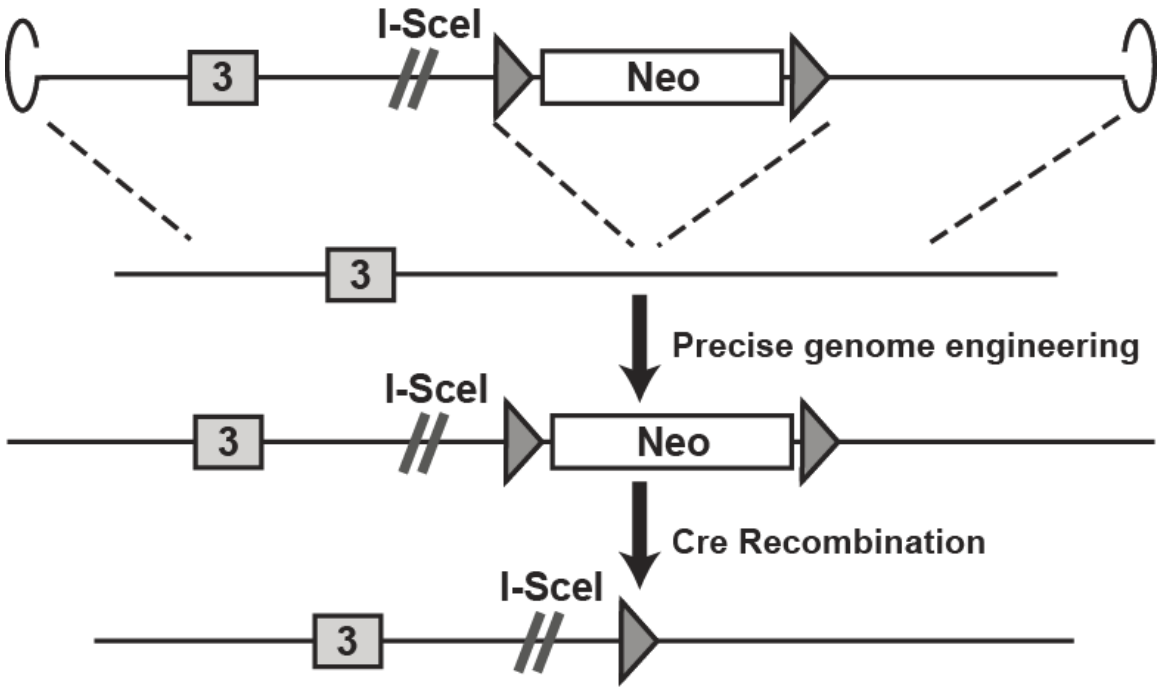

I-SceI
3
Neo
3
Precise genome engineering
I-SceI
3
Neo
Cre Recombination
I-SceI
3

Table S1. Conversion tracts of rAAV-mediated PGE in parental HCT116 cells

									Legends:	Viral	Genomic
NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI	#	%
+	+	+	+	+	+	+	-	-	-	5	2.2%
+	+	+	+	+	+	-	-	-	-	1	0.4%
+	+	+	+	+	-	+	-	-	+	1	0.4%
-	+	+	+	+	+	+	+	+	-	2	0.9%
-	+	+	+	+	+	+	+	-	-	14	6.1%
-	+	+	+	+	+	+	-	+	-	1	0.4%
-	+	+	+	+	+	+	-	-	-	9	3.9%
-	+	+	+	+	+	-	-	-	+	2	0.9%
-	+	+	+	+	+	-	-	-	-	17	7.4%
-	+	+	+	+	-	+	-	-	-	1	0.4%
-	+	+	+	+	-	-	-	-	-	7	3.0%
-	+	+	-	+	+	-	-	-	-	2	0.9%
-	+	-	+	+	+	+	-	-	-	2	0.9%
-	+	-	+	+	-	-	-	-	-	1	0.4%
-	+	-	-	-	+	+	+	+	-	1	0.4%
-	-	+	+	+	+	+	+	+	-	3	1.3%
-	-	+	+	+	+	+	+	-	-	3	1.3%
-	-	+	+	+	+	+	-	-	-	6	2.6%
-	-	+	+	+	+	-	+	-	-	2	0.9%
-	-	+	+	+	+	-	-	-	-	13	5.7%
-	-	+	+	+	-	-	-	-	-	4	1.7%
-	-	+	-	+	+	+	+	+	-	11	4.8%
-	-	+	-	+	+	+	+	-	-	5	2.2%
-	-	+	-	+	+	-	+	+	-	2	0.9%
-	-	+	-	+	+	-	-	-	-	1	0.4%
-	-	+	-	+	-	+	+	-	-	1	0.4%
-	-	+	-	+	-	-	+	-	-	1	0.4%
-	-	+	-	-	-	+	+	+	-	1	0.4%
-	-	-	+	+	+	+	+	+	+	2	0.9%
-	-	-	+	+	+	+	+	+	-	3	1.3%
-	-	-	+	+	+	+	+	-	-	1	0.4%
-	-	-	+	+	+	+	-	-	-	3	1.3%
-	-	-	+	+	+	-	+	-	-	1	0.4%
-	-	-	+	+	+	-	-	-	-	1	0.4%
-	-	-	-	+	+	+	+	+	+	5	2.2%
-	-	-	-	+	+	+	+	+	-	35	15.2%
-	-	-	-	+	+	+	+	-	-	16	7.0%
-	-	-	-	+	+	+	-	+	+	3	1.3%
-	-	-	-	+	+	+	-	+	-	2	0.9%
-	-	-	-	+	+	+	-	-	-	6	2.6%
-	-	-	-	+	+	-	+	-	-	2	0.9%
-	-	-	-	+	+	-	-	-	-	19	8.3%
-	-	-	-	+	-	+	+	+	-	2	0.9%
-	-	-	-	+	-	-	-	-	-	2	0.9%
-	-	-	-	-	+	+	+	+	-	3	1.3%
-	-	-	-	-	+	+	-	-	-	1	0.4%
-	-	-	-	-	+	-	-	-	-	4	1.7%
7	**66**	**115**	**105**	**220**	**209**	**148**	**116**	**76**	**13**	**230**	**100.00%**
3.04%	**28.70%**	**50.00%**	**45.65%**	**95.65%**	**90.87%**	**64.35%**	**50.43%**	**33.04%**	**5.65%**		

Table S2. Integration tracts of rAAV random integration in parental HCT116 cells

							Legends:	Viral	Genomic
NdeI	**EcoRI**	**LHP**	**NcoI**	**AseI**	**SspI**	**SacI**	**RHP**	**XbaI**	**SbfI**
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
37	**37**	**37**	**37**	**37**	**37**	**37**	**37**	**37**	**37**
100.00%	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

Table S3. Conversion tracts of dsDNA-mediated PGE in parental HCT116 cells

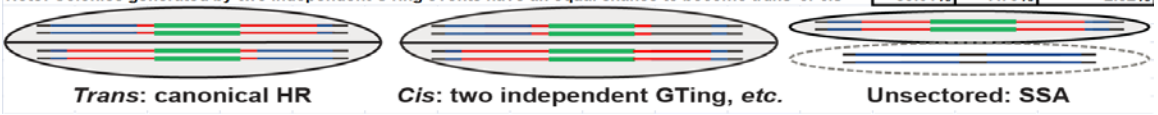
							Legends:	Viral	Genomic
NdeI	**EcoRI**	**LHP**	**NcoI**	**AseI**	**SspI**	**SacI**	**RHP**	**XbaI**	**SbfI**
+	+	+	+	+	+	-	-	-	-
-	+	+	+	+	+	+	-	+	-
-	+	+	+	+	+	+	-	+	-
-	+	+	+	+	+	-	-	-	-
-	+	-	-	+	+	+	-	-	-
-	-	+	+	+	+	+	+	+	+
-	-	+	+	+	+	+	+	+	-
-	-	+	+	+	+	+	-	-	-
-	-	+	+	+	+	-	-	-	-
-	-	+	+	+	+	-	-	-	-
-	-	+	+	+	-	+	-	-	-
-	-	-	+	+	+	+	+	-	-
-	-	-	-	+	+	+	+	+	+
-	-	-	-	+	+	+	+	-	-
-	-	-	-	+	+	+	-	-	-
-	-	-	-	-	+	+	+	+	-
-	-	-	-	-	+	+	-	+	-
-	-	-	-	-	-	-	-	-	-
1	**5**	**10**	**11**	**15**	**16**	**13**	**6**	**7**	**2**
5.56%	**27.78%**	**55.56%**	**61.11%**	**83.33%**	**88.89%**	**72.22%**	**33.33%**	**38.89%**	**11.11%**

Table S4. The sectoring assay in parental HCT116 cell

							Legends:	Viral	Sectored	Genomic
NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI	Sectored	Subcloned
+/-	+/-	+/-	+/-	+	+	+	+/-	+/-	-	2	#1
+/-	+/-	+/-	+	+	+	+	+/-	+/-	-	2	#2
+/-	+/-	+	+/-	+	+	+/-	+/-	+/-	+/-	2	#3
-	-	+	+/-	+	+	+/-	+/-	+/-	-	2	
-	+/-	+/-	+/-	+	+	+	+	+/-	-	2	#4
-	+/-	+/-	+/-	+	+	+	+	+/-	-	2	#5
-	+/-	+/-	+/-	+	+	+	+/-	+/-	-	2	#6
-	+/-	+/-	+/-	+	+	+/-	+/-	+/-	-	2	
-	+/-	+/-	+/-	+	+	+/-	+/-	+/-	-	2	#7
-	+/-	+/-	+/-	+	+	+/-	+/-	-	-	2	
-	+/-	+/-	+/-	+	+	+/-	-	-	-	2	#8
-	+/-	+/-	+/-	+	+/-	+/-	+/-	+/-	-	2	#9
-	-	+/-	+/-	+	+	+	+	+/-	+/-	2	
-	-	+/-	+/-	+	+	+	+	+/-	-	2	
-	-	+/-	+/-	+	+	+	+	+/-	-	2	
-	-	+/-	+/-	+	+	+/-	+/-	+/-	-	2	
-	-	+/-	+/-	+	+	+/-	-	+/-	-	2	
-	-	+/-	+/-	+	+/-	+	+	+/-	-	2	
-	-	-	+/-	+	+	+/-	+/-	+/-	-	2	#10
-	-	+/-	-	+	+	+/-	+	+	-	2	
-	+/-	+/-	+/-	+/-	+	+/-	+/-	+/-	-	2	#11
-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-	2	
-	-	+/-	+/-	+/-	+	+/-	+/-	+/-	-	2	
-	+/-	+	+	+	+	+	+	+	-	1	
-	+/-	+	+	+	+	+	+	-	-	1	
-	-	+	+/-	+	+	+	+	+	-	1	
-	+/-	+/-	+/-	+	+	+	-	-	-	1	
-	+/-	+/-	+/-	+	+	+	-	-	-	1	
-	+/-	+/-	+/-	+	+	-	-	-	-	1	
-	-	+/-	+/-	+	+	+	-	-	-	1	
-	-	-	+/-	+	+	+	+	+	-	1	
-	-	-	+/-	+	+	-	-	-	-	1	
-	-	+/-	-	+	+	+	-	-	-	1	
-	-	+/-	+/-	+/-	+	-	-	-	-	1	
-	+	+	+	+	+	+/-	+/-	-	-	1	
-	+	+	+	+	+	+/-	+/-	-	-	1	
-	-	+	+	+	+	+/-	+	+/-	-	1	
-	-	-	+	+	-	+/-	+/-	-	-	1	
-	-	-	-	+	+	+	+	+/-	-	1	
-	-	-	-	+	+	+	+	+/-	-	1	
-	-	-	-	+	+	+	+/-	+/-	-	1	
-	-	-	-	+	+	+	+/-	+/-	-	1	
-	-	-	-	+	+	+	+/-	-	-	1	
-	-	-	-	+	+	+/-	+/-	+/-	-	1	
-	-	-	-	+	+	+/-	+/-	+/-	-	1	
-	-	-	-	+	+	+/-	+/-	+/-	-	1	
-	-	-	-	+	+	+/-	-	+/-	-	1	
-	-	-	-	+	+/-	+/-	+	+/-	-	1	
-	-	-	-	-	+	+	+	+/-	-	1	
-	+	+	+	+	+	+	-	+	-	0	
-	+	+	+	+	-	-	-	-	-	0	
-	-	+	+	+	+	+	+	-	-	0	
-	-	+	+	+	+	-	+	-	-	0	
-	-	+	+	+	+	-	-	-	-	0	
-	-	-	+	+	+	+	+	-	-	0	
-	-	-	+	+	+	-	-	-	-	0	
-	-	-	-	+	+	+	+	+	-	0	
-	-	-	-	+	+	+	+	+	-	0	
-	-	-	-	+	+	+	+	-	-	0	
-	-	-	-	+	+	+	-	-	-	0	
-	-	-	-	+	+	+	-	-	-	0	
-	-	-	-	+	+	+	-	-	-	0	
-	-	-	-	+	+	-	-	-	-	0	
-	-	-	-	-	+	+	+	+	-	0	
-	-	-	-	-	+	+	+	+	-	0	
-	-	-	-	-	+	+	+	+	-	0	
3	18	26	29	4	4	23	23	31	2	49	
4.55%	27.27%	39.39%	43.94%	6.06%	6.06%	34.85%	34.85%	46.97%	3.03%	74.24%	

Table S5. Subcloning of the selected colonies from Table S4

												Legends:	Viral	Genomic	Unsectored
Clone	Subclone	NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI		*Trans*	*Cis*	Noncanonical
#1	1	-	-	-	-	+	+	+	+	+	-		1		
	2	+	+	+	+	+	+	+	-	-	-		1		
	3	+	+	+	+	+	+	+	-	-	-		1		
	4	-	-	-	-	+	+	+	+	+	-		1		
	5	+	+	+	+	+	+	+	-	-	-		1		
	6	+	+	+	+	+	+	+	-	-	-		1		
	7	+	+	+	+	+	+	+	-	-	-		1		
	8	-	-	-	-	+	+	+	+	+	-		1		
#2	1	+	+	+	+	+	+	+	-	-	-		1		
	2	-	-	-	+	+	+	+	+	+	-		1		
	3	-	-	-	+	+	+	+	+	+	-		1		
	4	-	-	-	+	+	+	+	+	+	-		1		
	5	-	-	-	+	+	+	+	+	+	-		1		
	6	+	+	+	+	+	+	+	-	-	-		1		
	7	+	+	+	+	+	+	+	-	-	-		1		
	8	+	+	+	+	+	+	+	-	-	-		1		
#3	1	-	-	+	-	+	+	+	+	+	+		1		
	2	+	+	+	+	+	+	-	-	-	-		1		
	3	-	-	+	-	+	+	+	+	+	+		1		
	4	-	-	+	-	+	+	+	+	+	+		1		
#4	1	-	-	-	-	+	+	+	+	+	-		1		
	2	-	+	+	+	+	+	+	+	-	-		1		
	3	-	+	+	+	+	+	+	+	-	-		1		
	4	-	-	-	-	+	+	+	+	+	-		1		
	5	-	-	-	-	+	+	+	+	+	-		1		
	6	-	-	-	-	+	+	+	+	+	-		1		
	7	-	-	-	-	+	+	+	+	+	-		1		
#5	1	-	-	-	-	+	+	+	+	+	-		1		
	2	-	+	+	+	+	+	+	+	-	-		1		
	3	-	-	-	-	+	+	+	+	+	-		1		
	4	-	+	+	+	+	+	+	+	+	-				1
	5	-	-	-	-	+	+	+	+	+	-		1		
	6	-	+	+	+	+	+	+	+	-	-		1		
	7	-	-	-	-	+	+	+	+	+	-		1		
	8	-	+	+	+	+	+	+	+	-	-		1		
#6	1	-	+	+	+	+	+	+	-	-	-		1		
	2	-	-	-	-	+	+	+	-	-	-			1	
	3	-	-	-	-	+	+	+	+	+	-		1		
	4	-	+	+	+	+	+	+	-	-	-		1		
	5	-	+	+	+	+	+	+	+	+	-			1	
	6	-	+	+	+	+	+	+	-	-	-		1		
	7	-	+	+	+	+	+	+	-	-	-		1		
	8	-	-	-	-	+	+	+	+	+	-		1		
#7	1	-	-	-	-	+	+	-	-	-	-			1	
	2	-	-	-	-	+	+	+	+	+	-		1		
	3	-	+	+	+	+	+	-	-	-	-		1		
	4	-	+	+	+	+	+	-	-	-	-		1		
	5	-	+	+	+	+	+	-	-	-	-		1		
	6	-	+	+	+	+	+	+	+	+	-			1	
#8	1	-	+	+	+	+	+	-	-	-	-		1		
	2	-	-	-	-	+	+	+	-	-	-		1		
	3	-	+	+	+	+	+	-	-	-	-		1		
	4	-	+	+	+	+	+	-	-	-	-		1		
	5	-	+	+	+	+	+	-	-	-	-		1		
	6	-	-	-	-	+	+	+	-	-	-		1		
	7	-	+	+	+	+	+	-	-	-	-		1		
	8	-	-	-	-	+	+	+	-	-	-		1		
#9	1	-	+	+	+	+	-	-	-	-	-		1		
	2	-	+	+	+	+	-	-	-	-	-		1		
	3	-	-	-	-	+	+	+	+	+	-		1		
	4	-	+	+	+	+	+	+	+	-	-				1
#10	1	-	-	-	-	+	+	+	+	+	-		1		
	2	-	-	-	+	+	+	-	-	-	-		1		
	3	-	-	-	+	+	+	-	-	-	-		1		
	4	-	-	-	-	+	+	+	+	+	-		1		
	5	-	-	-	-	+	+	+	+	+	-		1		
	6	-	-	-	-	+	+	+	+	+	-		1		
	7	-	-	-	-	+	+	+	+	+	-		1		
	8	-	-	-	-	+	+	+	+	+	-		1		
#11	1	-	+	+	+	+	+	-	-	-	-		1		
	2	-	+	+	+	+	+	-	-	-	-		1		
	3	-	+	+	+	+	+	-	-	-	-		1		
	4	-	-	-	-	-	+	+	+	+	-		1		
	5	-	-	-	-	-	+	+	+	+	-		1		
	6	-	-	-	-	-	+	-	-	-	-			1	
	7	-	-	-	-	-	+	-	-	-	-			1	
	8	-	+	+	+	+	+	-	-	-	-		1		
Note: Colonies generated by two independent GTing events have an equal chance to become *trans* or *cis*													89.61%	7.79%	2.82%



Trans: canonical HR *Cis*: two independent GTing, *etc.* Unsectored: SSA

Table S6. The sectoring assay in parental DLD-1 cells

						Legends:		Viral	Sectored	Genomic
NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI	Sectored
-	+/-	+	+	+	+	+/-	+	+	-	2
-	+/-	+	+	+	+	+/-	+	-	-	2
-	+/-	+	+	+	+	+/-	+/-	+/-	-	2
-	+/-	+	+	+	-	+/-	+/-	+/-	-	2
-	+/-	+	+/-	+	+	+	+	+/-	-	2
-	+/-	+	+/-	+	+	+	+/-	-	-	2
-	+/-	+	+/-	+	+	+/-	+	+	-	2
-	+/-	+	+/-	+	+	+/-	+/-	+/-	+/-	2
-	+/-	+	+/-	+	+	+/-	-	+	-	2
-	+/-	+	+/-	+	+	+/-	-	-	-	2
-	+/-	+	+/-	+	+	-	+	+/-	-	2
-	+/-	+/-	+/-	+	+	+/-	+/-	+	-	2
-	+/-	+/-	+/-	+	+	+/-	-	+/-	-	2
-	+/-	+/-	+/-	+	+	+/-	-	-	-	2
-	+/-	+/-	+/-	+	+	+/-	-	-	-	2
-	+/-	+/-	+/-	+	-	+/-	+/-	-	-	2
-	+/-	-	+	+	+	+/-	-	+	-	2
-	+/-	-	+/-	+	+	+/-	+	+	-	2
-	+/-	-	+/-	+	+	+/-	-	-	-	2
-	+/-	-	+/-	+	+/-	+	+	+	-	2
-	-	+/-	+/-	+	+/-	+/-	-	-	-	2
-	-	-	+/-	+	+	+	+	+/-	-	2
-	-	-	+/-	+	+	+	+	+/-	-	2
-	-	-	+/-	+	+	+/-	+	+/-	-	2
-	-	-	+/-	+	+	+/-	+/-	+/-	-	2
-	-	-	+/-	+	+	+/-	-	-	-	2
+/-	+	-	+	+	+	+	+	+	+	1
+/-	+/-	+	+	+	+	+	+	-	-	1
-	+/-	+	+/-	+	+	-	+	-	-	1
-	+/-	+/-	+/-	+	+	+	+	-	-	1
-	+/-	+/-	+/-	+	+	-	-	-	-	1
-	+/-	+/-	+/-	+	+	-	-	-	-	1
-	-	+/-	+/-	+	+	-	-	-	-	1
-	-	-	+/-	-	-	-	-	-	-	1
+	+	+	+	+	+	+	+	+/-	-	1
+	+	+	+	+	-	+/-	+	+/-	-	1
-	+	+	+	+	+	+/-	-	-	+/-	1
-	+	+	+	+	+	+/-	-	-	-	1
-	-	+	+	+	+	+/-	+/-	-	-	1
-	-	+	+	+	+	+/-	-	-	-	1
-	-	-	-	+	+	+/-	-	-	-	1
-	+	+	+	+	+	+	-	-	-	0
-	-	+	+	+	+	+	-	-	-	0
-	-	+	-	+	+	+	+	-	-	0
-	-	-	-	+	+	+	+	+	-	0
-	-	-	-	-	+	+	-	-	-	0
2	25	10*	27	0	2	25	8	12	2	41
4.35%	54.35%	21.74%	58.70%	0.00%	4.35%	54.35%	17.39%	26.09%	4.35%	89.13%

Table S7. Conversion tracts of I-SceI-induced rAAV PGE on the right arm

							Legends:	Viral	Genomic
NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI	Half *I-Sce*
-	-	-	-	+	+	+	+	-	-	
-	+	+	+	+	+	+	+	+	+	
-	-	-	-	+	+	+	+	-	-	
-	-	-	-	-	+	+	+	+	-	
+	+	+	+	+	+	+	-	-	-	1
-	-	-	+	+	+	+	+	-	-	
-	+	+	+	+	+	+	+	-	-	
-	-	+	+	+	+	+	+	+	-	
-	-	-	-	-	+	+	+	+	-	
+	+	+	+	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	+	-	
-	+	+	+	+	+	+	-	-	-	1
-	-	-	-	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	-	-	
-	+	+	+	+	+	+	+	+	+	
+	+	-	+	+	+	+	+	+	-	
-	-	-	-	+	+	+	-	-	-	1
-	-	-	-	+	+	+	+	-	-	
-	-	-	-	+	+	+	+	-	-	
-	-	-	-	-	+	+	+	+	-	
-	-	-	-	+	+	+	+	+	-	
-	+	+	+	+	+	+	-	-	-	1
-	-	-	-	+	+	+	+	+	-	
-	-	-	-	+	+	+	-	-	-	1
-	-	-	+	+	+	+	+	-	-	
-	-	-	-	+	+	+	+	+	-	
-	-	+	-	+	+	+	+	+	-	
-	-	-	+	+	+	+	+	+	-	
+	+	+	+	+	+	+	+	-	-	
-	-	-	-	+	+	+	+	-	-	
-	+	+	+	+	+	+	+	-	-	
+	+	+	+	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	+	-	
-	-	+	+	+	+	+	+	+	-	
-	+	-	-	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	+	-	
-	+	+	+	+	+	+	+	+	-	
-	+	+	+	+	+	+	+	-	-	
-	-	+	+	+	+	+	+	-	-	
-	-	-	-	-	+	+	+	-	-	
-	-	+	+	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	-	-	
-	-	+	+	+	+	+	+	+	-	
-	-	-	-	+	+	+	-	-	-	1
-	-	-	+	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	+	-	
-	-	+	-	+	+	+	+	+	-	
-	-	+	-	+	+	+	+	-	-	
-	+	+	+	+	+	+	+	-	-	
+	+	+	+	+	+	+	-	-	-	1
-	-	-	-	+	+	+	-	-	-	1
-	-	-	-	+	+	+	+	+	+	
-	-	+	+	+	+	+	+	-	-	
-	-	+	+	+	+	+	+	-	-	
-	-	-	-	+	+	+	+	-	-	
-	+	+	+	+	+	+	+	-	-	
-	-	-	-	+	+	+	-	-	-	1
-	-	-	-	-	+	+	-	-	-	1
-	+	+	+	+	+	+	-	-	-	1
-	+	+	+	+	+	+	+	-	-	
-	-	-	-	+	+	+	+	+	-	
-	-	-	-	+	+	+	+	-	-	
-	-	-	-	+	+	+	+	+	-	
6	19	27	29	59	64	64	53	30	3	11
9.38%	29.69%	42.19%	45.31%	92.19%	100%	100%	82.81%	46.88%	4.69%	17.19%

Table S8. Conversion tracts of I-SceI-induced rAAV PGE on the left arm

							Legends:	Viral	Genomic
NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI
-	-	+	+	+	+	+	ND	+	+
-	-	+	+	+	+	-	ND	-	-
-	+	+	+	+	+	+	ND	+	-
-	-	+	+	+	+	+	ND	-	-
-	-	-	+	+	+	+	ND	+	+
-	-	-	+	+	+	+	ND	+	-
-	-	+	+	+	+	-	ND	-	-
-	-	-	+	+	+	+	ND	-	-
-	+	+	+	+	+	+	ND	-	-
-	+	+	+	+	+	+	ND	-	-
-	+	+	+	+	+	+	ND	+	-
-	+	+	+	+	+	+	ND	-	-
+	+	+	+	+	+	+	ND	-	-
-	-	+	+	+	+	+	ND	+	+
-	-	-	+	+	+	-	ND	-	-
-	+	+	+	+	+	-	ND	-	-
-	-	-	+	+	+	+	ND	+	+
-	-	-	+	+	+	+	ND	+	-
+	+	+	+	+	+	+	ND	+	+
-	+	+	+	+	+	+	ND	+	-
-	-	-	+	+	+	+	ND	+	-
-	-	-	+	+	+	+	ND	-	-
-	-	+	+	+	+	+	ND	-	-
-	+	+	+	+	+	-	ND	-	-
-	-	-	+	+	+	+	ND	+	-
-	+	+	+	+	+	+	ND	+	-
-	+	+	+	+	+	+	ND	-	-
-	+	+	+	+	+	-	ND	-	-
-	-	-	+	+	+	+	ND	-	-
-	+	+	+	+	+	+	ND	+	-
-	+	+	+	+	+	+	ND	+	-
-	+	+	+	+	-	-	ND	-	-
-	-	+	+	+	+	-	ND	-	-
-	-	+	+	+	+	-	ND	+	-
+	+	+	+	+	+	+	ND	-	-
-	-	+	+	+	+	+	ND	+	-
-	-	+	+	+	+	+	ND	-	-
-	-	-	+	+	+	+	ND	+	-
-	-	+	+	+	-	-	ND	-	-
-	-	+	+	+	+	+	ND	-	-
+	+	+	+	+	+	+	ND	-	-
-	-	+	+	+	-	-	ND	-	-
-	-	+	+	+	+	+	ND	+	-
+	-	+	+	+	+	-	ND	-	-
+	+	+	+	+	+	+	ND	-	-
-	-	-	+	+	+	+	ND	+	-
-	-	-	+	+	+	+	ND	-	-
-	-	+	+	+	+	+	ND	-	-
6	19	35	48	48	45	36	ND	20	5
12.50%	39.58%	72.92%	100.00%	100.00%	93.75%	75.00%	ND	41.67%	10.42%

Table S9. Conversion tracts of rAAV-mediated PGE colonies in MLH1+ cells

						Legends:		Viral	Genomic
NdeI	**EcoRI**	**LHP**	**NcoI**	**AseI**	**SspI**	**SacI**	**RHP**	**XbaI**	**SbfI**
+	+	+	+	+	+	+	-	-	-
-	+	+	+	+	+	+	+	+	-
-	+	+	+	+	+	+	+	-	-
-	+	+	+	+	+	+	+	-	-
-	+	+	+	+	+	-	-	-	-
-	-	+	+	+	+	+	+	-	-
-	-	+	+	+	+	+	-	+	-
-	-	+	+	+	+	-	+	-	-
-	-	+	+	+	+	-	-	-	-
-	-	+	+	+	-	-	+	-	-
-	-	-	+	+	+	+	-	-	-
-	-	-	-	+	+	+	+	+	-
-	-	-	-	+	+	+	+	+	-
-	-	-	-	+	+	+	+	+	-
-	-	-	-	+	+	+	-	+	-
-	-	-	-	+	+	-	-	-	-
-	-	-	-	+	+	-	-	-	-
-	-	-	-	+	+	-	-	-	-
-	-	-	-	+	-	+	-	-	-
-	-	-	-	+	-	-	-	-	-
1	**5**	**10**	**11**	**20**	**17**	**12**	**9**	**6**	**0**
5.00%	**25.00%**	**50.00%**	**55.00%**	**100.00%**	**85.00%**	**60.00%**	**45.00%**	**30.00%**	**0.00%**

Table S10. Integration tracts of rAAV random integration in MLH1+ cells

Legends: Viral | Genomic

NdeI	EcoRI	LHP	NcoI	AseI	SspI	SacI	RHP	XbaI	SbfI
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
21	21	21	21	21	21	21	21	21	21
100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Materials and Methods

Cell culture

The HCT116 and DLD-1 cell lines were cultured in McCoy's 5A medium supplemented with FBS, L-glutamine, penicillin and streptomycin with 5% CO_2 at 37 ℃.

Cell lines and plasmids

The HCT116 cell line was obtained from ATCC. The $MLH1^+$ cell line was provided by Horizon Discovery, Ltd. The DLD-1 cell line was obtained from Dr. D. Largaespada. The DR-GFP and SA-GFP reporter plasmids were obtained from Dr. M. Jasin and the Rad51K133A expression vector was obtained from Dr. J. Stark [134].

Viruses

Briefly, the left and right homology arms were amplified by PCR from HCT116 genomic DNA. Viral SNPs were introduced using a QuickChange™ site-directed mutagenesis kit. The arms were then joined with a drug selection cassette using fusion PCR and the resulting product was ligated to a pAAV backbone. All virus packaging and infections were performed as described [129].

Conversion tracts analysis

Genomic DNA was Isolated and the homology arms of the GT and RI clones were amplified by diagnostic PCRs (Figure 1C). The retention of the vector-borne markers was analyzed first by restriction enzyme digestion and then confirmed by sequencing.

Repair assays

Briefly, cells were subcultured in 6-well tissue culture plates. The next day, the cells were transfected with 0.5 μg mCherry, 1.0 μg of an *I-SceI* expression plasmid and

1.0 μg DR-GFP or SA-GFP assay substrates. GFP and mCherry expression was then analyzed 48 hr post transfection using flow cytometry. The repair efficiency was calculated as the percentage of GFP and mCherry doubly positive cells divided by the mCherry-positive cells. For the Rad51DN experiment, an additional 1.0 μg of the Rad51K133A expression plasmid was transfected as well.

The targeting efficiency assay at the HPRT locus

Briefly, cells were subcultured in 6-well tissue culture plates on day 1. On day 2, 100 μl of the appropriate viral stock was added to the wells. On day 4, the cells were counted and aliquoted into 10 cm tissue culture dishes for drug selection. The plates were supplemented with 1 mg/ml G418 or 0.5 mg/ml G418 plus 5 μg/ml 6-thioguanine for 12 days. The PGE and random integration efficiencies were calculated as the number of G418-resistant 6-thioguanine-resistant and G418-resistant 6-thioguanine-sensitive colonies per 10^6 cells, respectively. Results were averaged from 7 plates. For the Rad51DN experiment, cells were transfected with 2.5 μg Rad51K133A expression plasmid 48 hr before infection. For the *I-SceI* experiments, cells were transfected with 2.5 μg of an *I-SceI* expression plasmid 12 or 20 hr before infection.

Ligation-mediated PCR

Genomic DNA was isolated at the designated times after *I-SceI* induction. DNA (1 μg) was ligated with 100 pmol of adaptors at 16℃ overnight. PCR was performed at the linear stage using a 25 ng ligation product with the primers illustrated in Figure 5D. β-actin primers were used as loading control.

CHAPTER III: THE MECHANISMS OF PRECISE GENOME ENGINEERING USING ODN DONORS

Summary

The use of programmable meganucleases is quickly transforming genome editing and functional genomics. CRISPR/Cas9 was developed such that targeted genomic lesions could be introduced in vivo with unprecedented ease. In the presence of homology donors, these lesions facilitate high-efficiency precise genome editing (PGE) via the homology-directed repair (HDR) pathways. However, the identity, property and hierarchy of the HDR pathways leading to the formation of PGE products remain elusive.

Here we established an EGFP > BFP conversion system to systematically characterize oligonucleotide (ODN)-mediated PGE using Cas9 and its nickase variants in human cells. We demonstrate that, unlike dsDNA donors with central heterologies, ODNs generated short conversion tracts with Gaussian distributions. Interestingly, single nick-induced PGE using ODN donors produced either unidirectional or bidirectional conversion tracts depending on the relative strandedness of the ODNs and the nick. Moreover, the ODNs were physically incorporated into the genome only in the bidirectional, but not in the unidirectional, conversion pathway. In the presence of compound genomic lesions, the unidirectional conversion pathway was preferentially utilized even though the knock-in mutation could theoretically be converted by both pathways. Collectively, our results demonstrate that ODN-mediated PGE utilizes the synthesis-dependent strand annealing and single-stranded DNA incorporation pathways. These pathways generate short conversion tracts with Gaussian distributions. Although synthesis-dependent strand annealing is preferentially utilized, our work unequivocally establishes the existence of a robust single-stranded DNA incorporation pathway in human cells. This work extends the paradigms of HDR-mediated gene conversion and establishes practical guidelines for PGE in human cells.

Introduction

Genome editing is the intentional alteration of the genetic information in living cells or organisms. Since Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR-associated 9 (CRISPR/Cas9) was repurposed for genome editing [56-60], the speed of CRISPR methodology is quickly bridging the genotype and phenotype worlds and facilitating high-throughput reverse genetic studies [94-99]. In the CRISPR age, targeted genomic lesions can be introduced with unprecedented ease, which facilitate either high-efficiency gene disruption by non-homologous end joining (NHEJ) or, in the presence of donor DNA, precise genome editing (PGE) by homology-directed repair (HDR) [12]. Research models of the desired genotype can be expeditiously generated within one cell cycle of most genetically-tractable organisms [100-103], and within one round of subcloning of cultured cells [167, 168]. Although HDR is the more desirable pathway for genome editing and gene therapy, its activity is much lower than NHEJ and highly dependent on the cell cycle status in human cells [8, 16, 50]. Thus, marker-free PGE still requires the laborious screening of hundreds of subclones, especially in hard-to-transfect cells or those with low HDR activity or limited tolerance of genomic lesions. Moreover, bi-allelic and multiplexed PGE is still challenging in human cells.

Complicating our understanding, HDR is composed of at least six different pathways in eukaryotic cells [7, 21]. Four of these pathways can lead to PGE without significantly sacrificing genomic stability - the double-strand break repair (DSBR) model [22, 23], Holliday junction dissolution [26-28], synthesis-dependent strand annealing (SDSA) [31-33] and single-strand DNA incorporation (ssDI, also sometimes referred to as single-strand assimilation) [38-40] pathways **(Figure S1)**. PGE is defined by the

fraction of HDR events leading to the conversion of desired knock-in mutations using exogenous homology donors. Importantly, the identity, property and hierarchy of usage of the HDR pathways leading to the formation of PGE products remains a knowledge barrier that hinders the improvement of PGE.

In the previous chapter, we demonstrated that in human somatic cells the introduction of large knock-in mutations with dsDNA donors (including transfected plasmids or viruses with dsDNA intermediates) must engage the DSBR pathway for PGE [169]. Interestingly however, many laboratories have reported that single-stranded oligonucleotides (ODNs) can be used to also obtain robust PGE [72-76]. This is surprising since ODN donors are likely physically unable to form Holliday junctions and therefore cannot engage DSBR. Instead they must perforce utilize some form of the SDSA or ssDI pathways during PGE [39, 170]. In the absence of a targeted meganuclease like Cas9, the ODN donors likely must use spontaneously occurring genomic lesions in order to perform PGE. Indeed, it was proposed that ODNs could be incorporated into the gaps generated by nucleotide excision repair [72, 73] or lagging strand synthesis during DNA replication [74-76], although the specific mechanisms are still disputed [77, 78]. Interestingly, both of these models involve the physical incorporation of the ODNs into the genome via an ssDI-like mechanism; a prediction that was later supported by experiments carried out by Radecke *et al* [38]. Confusingly, however, the same group of authors also demonstrated that ODNs could not be detected being physically incorporated into the genome during double-strand break (DSB)-induced PGE, suggesting the existence of an alternative mechanism(s) [79]. Finally, in the presence of meganuclease-induced targeted single-strand nicks, it was proposed by Davis and Maizels that certain forms of the SDSA and the ssDI pathways could be utilized depending on the strandedness of the donor ODNs [40, 80]. However,

the exact mechanisms, conversion tracts and hierarchy of these pathways have remained elusive.

Here, we have systematically dissected the molecular mechanisms of ODN-mediated PGE induced by Cas9, Cas9 D10A nickase (nCas9) and paired-nickases using two lines of evidence: the directionality of the conversion tracts and the physical incorporation of the ODN donors. We demonstrate that ODN-mediated PGE utilizes the SDSA and ssDI pathways. In contrast to DSBR, these pathways generate short conversion tracts with one-sided and two-sided Gaussian distributions, respectively. In the presence of compound genomic lesions such as DSBs and paired nicks, the SDSA pathway is preferentially utilized.

Results

The EGFP > BFP conversion system

To systematically investigate the molecular mechanism of meganuclease-induced PGE using ODN donors, we established the EGFP > BFP conversion system **(Figure 1A)**. This system takes advantage of the sequence similarity of EGFP and BFP. By introducing T65S and Y66H mutations into the chromophore domain, EGFP can be completely converted into BFP [171]. To this end, using recombinant adeno-associated virus (rAAV)-mediated PGE [129, 169], we introduced the CMV-EGFP-pA expression cassette into the HPRT locus on the X chromosome of human HCT116 cells, in both sense (S) and antisense (AS) directions with respect to the direction of the HPRT gene **(Figures S2A and S2B)**. Targeted integration in either direction results in mono-allelic EGFP expression and the inactivation of the HPRT gene. After rAAV infection, the HPRT-negative cells were enriched using 6-thioguanine selection, and individual subclones with bright EGFP expression were screened using designated primers **(Figures S2A and S2B).** The EGFP cassette was precisely knocked into the HPRT gene in 1 out of 4 subclones in the sense orientation and 2 out of 4 subclones in the antisense orientation **(Figures S2C and S2D)**. One individual subclone from each orientation was designated as HPRT-EGFP sense and antisense cell lines, respectively, and used for subsequent studies. Targeted genomic lesions such as nicks and DSBs were subsequently introduced into these reporter cell lines within the chromophore sequence using Cas9 or nCas9 from *S. pyogenes* and sgRNAs in both sense (S) and antisense (AS) orientations with respect to EGFP **(Figure 1A)**. In the presence of ODN donors bearing the SH mutations (ACCTAC > TCTCAT), EGFP could efficiently be

converted into BFP, which generated distinguishable spectra under flow cytometry **(Figure 1C and Figure S3B)**.

The efficiency of ODN-mediated PGE

We first determined the efficiency of ODN-mediated PGE using nCas9 with ODNs (BFP_S160 and BFP_AS160, see **Supplemental Sequences** for more details) and sgRNAs in both S and AS orientations with respect to EGFP **(Supplemental Sequences)**. By transfecting the HPRT-EGFP cell lines with the nCas9 and sgRNA expression vectors and ODN donors all in a DNA format, we could routinely induce efficient EGFP to BFP conversions as quantitated using flow cytometry **(Figure 1 and Figure S3)**. In contrast, the omission of the donor ODNs resulted in <0.01% conversion into blue cells (data not shown). Interestingly (see the Discussion for additional discussion of this topic), the conversion frequency in the HPRT-EGFP antisense cell line was routinely 2 to 4 times higher than compared to the conversion frequency in the HPRT-EGFP sense cell line **(Compare Figure 1 with Figure S3)**. Because the HPRT-EGFP antisense cell line consistently generated a higher BFP conversion efficiency, this cell line was chosen for subsequent studies. In contrast to the bias observed with HPRT-EGFP orientation, the difference in conversion efficiency between any of the four possible sgRNA and ODN pair configurations was trivial with an average of ~1.5% conversion **(Figure 1B, white bars).** Thus, no significant strand bias for either the nick or the ODN donor was observed and this was also true for the HPRT-EGFP sense cell line **(Figure S3A)**.

We next examined the efficiency of ODN-mediated PGE in the presence of double-strand genomic lesions in the HPRT-EGFP antisense cell line. Wildtype Cas9 cleavase and nCas9 paired-nickases [172-174] generate double-stranded blunt or sticky

overhangs, respectively. Depending on the orientations of the protospacer adjacent motif (PAM) sequences, paired-nickases are generally classified into PAM-out and PAM-in configurations, which generate 5' and 3' overhangs, respectively. Impressively, all of the double-stranded genomic lesions induced PGE with consistently higher efficiency (~2-fold) than single nicks **(Figure 1B, compare filled bars with white bars)**. The only exception to this observation was in the case of the dual nickases utilizing PAM-in sgRNAs and AS donor ODNs, which had a conversion frequency more comparable to that induced by single nicks. As we discuss below, this configuration is indeed a special case. In summary, these experiments demonstrated that both nicks and (even more so) double-stranded lesions can induce robust PGE in human somatic cells by utilizing donor ODNs.

The mechanisms of single-nick-induced PGE using complementary strand ODNs

To map the associated conversion tracts, some of the ODNs were labeled with three-to-a-side synonymous single-nucleotide polymorphisms (SNPs) on both sides of the TY > SH mutations in either a distributive or clustered fashion (6SNPs_A and SNPs_B, respectively) in both S and AS orientations **(Figures 1A, 2F and Supplemental Sequences**). The BFP-positive cells were enriched using FACS sorting and single-cell subcloned via limiting dilution. The BFP conversion of the individual subclones was subsequently confirmed by PCR using mismatch sensitive primers (BFP_CF and BFP_ER, **Supplemental Sequences**). The individual conversion tracts of the BFP-positive subclones were then amplified using flanking primers (BFP_EF and BFP_ER, **Supplemental Sequences**) and sequenced using Sanger sequencing **(Tables S1-9)**. The SNP retention curves were compiled by plotting the frequency of SNP retention against the distance of each SNP to the center of the SH mutations. The

retention frequency of SNPs on both the 6SNP_A and 6SNP_B ODNs in the same configuration were overlaid into a single SNP retention curve **(Figures 2 and 4)**.

In contrast to dsDNA donors containing a long central heterology [169], ODNs-mediated PGE produced short conversion tracts with Gaussian distributions **(Figures 2A-E)**. When PGE was initiated by nCas9 and a single sgRNA, ODNs complementary to the strand with the nick produced conversion tracts with a ~one-sided Gaussian distribution biased in the 3' direction relative to the nick **(Figure 2A)**. Impressively, the distribution of the conversion tracts was almost perfectly mirrored when the strandedness of the nick and the ODNs were simultaneously inverted **(Figure 2B and Table S2)**. In both cases, the SNP retention curve was composed of a mini plateau (very high co-conversion region) near the SH mutations being selected for, followed by a sigmoidal decay in the 3' direction, and a steep drop-off in the 5' direction. For example, CAC, as the first SNP on the 6SNP_A ODN 3' to the nick, was retained in 46.5% of the conversion tracts, whereas GTC, its counterpart 5' to the nick, was converted a mere 2.3% of the time **(Figure 2A and Table S1)**. The corresponding SNPs on the 6SNP_B ODNs were converted with a similar asymmetry, although not always as dramatically as the ones on 6SNP_A. This was probably due to the fact that the SH mutations and the proximal SNPs were clustered tightly together and therefore were frequently recognized as one piece of heterology during gene conversion. The co-conversion of this larger heterology perforce utilized more distal homologies and generated longer conversion tracts.

The conversion tracts generated by ODNs complementary to the strand with the nicks **(Figure 2A, 2B)** were consistent with an SDSA model [31-33], in which gene conversion stems from unidirectional DNA synthesis using the ODN sequence as a template [40] **(Figure 3A, top)**. After a genomic lesion such as a single-strand nick is

introduced, the 3' strand of the lesion base pairs with the homology region on a complementary-strand ODN during homology search. Then the invading strand copies the downstream genetic information from the ODN donor via 5' to 3' DNA synthesis. Interestingly, if a small part of the selected knock-in mutations are upstream (*i.e*., 5') to the genomic lesion (as 4 or 5 nts are in our system), the 3'-end of the invading strand needs to be resected to accommodate the conversion of the heterology region before DNA synthesis begins. Because we need to select for the conversion of the SH mutations and use proximal SNPs to map the conversion tracts nearby (which favors co-conversion) in our experimental strategy, it is hard to address how often this 3' end processing occurs in natural HDR where these constraints would not apply. Whatever the frequency, however, it nevertheless seems that the 3' end processing rarely goes beyond 10 bp **(Figures 2A, 2B and 3A)**. Then DNA synthesis proceeds past the knock-in mutations before the invading strand anneals back with its original partner. If DNA synthesis travels beyond the genomic lesion, a flap will be created after annealing. The flap can be cleaved by structure-specific endonucleases and the resultant gap filled by DNA synthesis and ligation as part of the standard flap metabolism [175]. In summary, part of the ODN sequence including the knock-in mutations will be converted as a result of DNA synthesis starting from up to 10 bp 5' of the genomic lesion **(Figures 2A, 2B and 3A, bottom)**. Since the synapse formed by strand invasion cannot be stabilized by the formation of double Holliday junctions, the invading strand falls back quickly after DNA synthesis bridges the genomic lesion [21], generating short conversion tracts with an average of 20 bp in the 3' direction **(Figures 2A, 2B and 3A, bottom)**.

The mechanisms of single-nick-induced PGE using same strand ODNs

In contrast to ODN donors corresponding to the complementary strand, ODN donors corresponding to the same strand that was nicked produced short conversion tracts with more symmetrical two-sided Gaussian distributions **(Figures 2C and 2D)**. For example, CAC, the first SNP in the 5' direction on the 6SNP_A ODN, was retained 20.5% of the time as compared to 31.8% for GTC, the corresponding first SNP in the 5' direction **(Figure 2C and Table S3)**. Again, this SNP retention curve was mirrored when the strandedness of the nick and the ODNs were simultaneously inverted **(Figure 2D and Table S4)**. We also noted a modest, but consistent, 5' bias in these SNP retention curves, in which the SNPs 5' to the nick were consistently converted with higher frequency than their downstream counterparts **(Figures 2C and 2D)**. Interestingly, this bias is opposite to the direction of common exonuclease resection [176].

The bidirectional conversion tracts generated by ODNs same to the strand with the nicks were consistent with an ssDI model, in which the ODN is physically assimilated into a single-strand gap and displaces the flanking sequences on both sides [38-40] (**Figure 3B)**. Although various forms of the ssDI model have been previously proposed [72-76], the efficiency of this pathway was so low in the absence of targeted genomic lesions that it was generally not regarded as a naturally-occurring HDR pathway (also note that ssDI is different from the more familiar process of single-strand annealing; SSA) [7, 21]. Interestingly, it was suggested that transcript-RNA-mediated HDR occurs via a mechanism similar to ssDI, although the RNA donor has to fall off after it bridges the complementary strand genomic lesion [44]. In the presence of targeted genomic lesions, it was also not known how the lesions were processed into the single-stranded gaps required for ssDI. In our model **(Figure 3B)**, we propose that the initial gap is generated

by futile homology search when the 3'-end is trying to engage in the strand invasion *ala* the SDSA pathway **(Figure 3A)**. When the ODN is initially assimilated into the gap generated by a futile homology search, it will produce the subtle 5' asymmetry/bias that was actually observed in our SNP retention curves **(Figures 2C and 2D)**. Bidirectional conversion tracts are then generated as the ODN displaces the flanking sequences via a branch migration-like activity [159], while 5' to 3' exonuclease resection may also contribute to this process [176]. Eventually, the flaps of the ODN and genomic DNA are processed by flap endonucleases [175], and the ODN are physically incorporated into the chromosome in both directions from the genomic lesions with a bias towards the 5' direction from the lesion.

The mechanism of DSB-induced PGE using ODN donors

Double-stranded genomic lesions can theoretically engage both the SDSA and ssDI pathways in the presence of ODN donors. To elucidate the hierarchy of HDR pathways utilized in the presence of double-strand genomic lesions, we first compiled the SNP retention curve of ODN-mediated PGE induced by wildtype Cas9 cleavase. When a DSB was introduced by Cas9, it generated a conversion tract with a ~one-sided Gaussian distribution **(Figure 2E)** that was virtually indistinguishable from the conversion track generated by a single-strand nick in a strand complementary to the ODN **(compare Figure 2E, Table S5 with Figure 2A, Table S1).** The ~unidirectional conversion tracts indicated that, in the presence of a blunt-ended DSB directly at the position of the knock-in mutation, PGE is preferentially mediated by the SDSA pathway using the lesion complementary to the strand with the ODNs **(Figure 3A),** even though the lesion can likely initiate effective ssDI as well. This result nicely explains a previous paradox that the physical incorporation of ODN donors was detected in the absence [38],

but not presence [79], of targeted chromosomal DSBs: without a targeted lesion, the ODNs were likely physically incorporated via the ssDI pathways [72-76] using naturally occurring gaps, whereas with a DSB most of the PGE products are generated via the SDSA pathway as a result of DNA synthesis using ODNs as templates **(Figures 2E and 3A)**.

The mechanism of paired-nick-induced PGE using ODN donors

The use of nCas9 in conjunction with dual sgRNAs in PAM-out and PAM-in configurations generates paired nicks with 5' or 3' overhangs, respectively [172-174]. These overhangs can theoretically engage the ODNs via both the SDSA and ssDI pathways. In addition, since paired nicks can be effectively placed up to 60 bp away from each other, it was not known whether they are processed into double-strand gaps. To illustrate the mechanisms of paired-nick-induced PGE, we compiled the SNP retention curves using nCas9 paired-nickases in both PAM-out and PAM-in configurations in the presence of S and AS ODN donors. We discovered that these paired-nickases generated conversion tracts with much larger plateaus **(Figure 4)**, compared to those of the respective single nicks **(Figure 2)**. The SNPs between the paired nicks were retained with an average of more than 75% frequency **(Tables S6, S7)**. Besides the plateau, the PAM-out paired-nickases generated a one-sided SNP retention curve with a ~sigmoidal decay on the left, and a much steeper drop-off on the right **(Figure 4A)**, whereas the PAM-in configuration with a S ODN donor generated two ~sigmoidal curves **(Figure 4B)** that were farther apart from each other, compared to the SNP retention curve of the corresponding single nickase **(Figure 2D)**.

We propose that the 5'-overhangs generated by PAM-out paired-nickases can be processed by extensive exonuclease resection [176] to generate large double-stranded

gaps **(Figure 5A)** that ultimately yield the broad plateaued SNP retention curve **(Figure 4A)**. We believe that, following this resection, PGE is likely initiated by strand invasion of a now uncovered 3'-end complementary to the strand of the ODNs via the SDSA pathway **(Figure 5A)**. In the PGE products, DNA synthesis would have to travel a long distance to bridge the extended double-strand gaps. After annealing with the 3' end on the other side of the lesion, the resultant flap on the AS strand is processed and single-strand gap on the S strand is filled in by standard gap repair. We believe that the steep drop-off on the right side of the SNP retention curve **(Figure 4A)** is an SDSA SNP retention signature **(Figure 2A, 2B)** indicative of the minimal 3' end processing before the start of DNA synthesis, and that strand invasion is initiated by the 3' overhang on the AS strand **(Figure 5A)**. The broad plateaued conversion tracts **(Figure 4A)** are thus evidence that the 5' overhangs of the PAM-out paired-nickases are processed into double-strand gaps, so that the subsequent DNA synthesis has to travel beyond these gaps to repair the genomic lesions in the PGE products **(Figure 5A)**.

In contrast, in the PAM-in configuration, the 3'-overhangs initially generated by paired nicking are not subject to extensive exonuclease resection [176] **(Figure 5B)**. Because the 3' overhang on the AS strand complementary to the ODNs cannot effectively convert the knock-in mutations (unless it is resected by > 30 nts), the majority of the PGE products must be generated by the ssDI pathway. We propose that the presence of AS strand nicks may loosen the 3'-ends on the chromosomal S strand and mobilize it for a (in this instance, futile) homology search **(Figure 5B)**. This would provide a window of opportunity for the S strand ODN to anneal to the gap generated by futile homology search before further bidirectional strand displacement occurs. In total, this would produce the longer conversion tracts observed compared to those of single-nick-induced PGE via the ssDI pathway (**compare Figures 2D and 4B**).

The hierarchy of ODN-mediated PGE in the presence of compound genomic lesions

For the paired-nickase-induced PGE using S strand ODNs, we notice that the knock-in mutations and 3' overhang on the AS strand are close to each other in the PAM-out configuration, but far away in the PAM-in configuration **(Figures 4A and 4B)**. Even if SDSA is the preferred pathway for double-strand lesions **(Figure 2E)**, it cannot effectively convert the knock-in mutations using the AS strand overhang in the PAM-in configuration, unless extensive resection of the 3'-end occurs **(Figure 4B)**. To dissect whether the hierarchy of paired-nickases-induced PGE is determined by the PAM configuration or the relative position of the knock-in mutation and the 3' overhang on the complementary strand, we also compiled the SNP retention curves of paired-nickases-induced PGE using AS strand ODNs **(Figures 4C and 4D)**.

PAM-out paired-nickases with AS strand ODNs produced a compound SNP retention curve with a plateau, a sigmoidal decay on the right and a linear slope on the left side **(Figure 4C and Table S8)**. We inferred that the conversion tracts were predominantly generated by ssDI, because SDSA would have produced continuous conversion tracts downstream of the 3' overhang on the S strand (and thus ACC, TCC and CAC would have been retained with 100% frequency — which was not observed). We thus propose that the 5' overhangs generated by PAM-out paired-nickases are first resected by exonucleases [176] to expose a single-strand gap for ssDI **(Figure 6A)**. Since resection on both strands may occur to a random extent before an incoming ODN is assimilated to the 5' overhang on the S strand, the left border of the single-strand gaps should be randomly distributed within a certain region before the initial ssDI occurs. Subsequently, strand displacement occurs in both directions, which creates the sigmoid

decay on the right of the SNP retention curve. At the same time, the linear slope on the left is shaped by the accumulation of individual ssDI events with random left borders **(Figures 4C and 6A, bottom)**.

On the other hand, PAM-in paired-nickases generated a two-sided SNP retention curve **(Figure 4D)** similar to that of single-nick-induced ssDI **(Figure 2C)**. We propose that the bulk of the PGE products are generated via a simple SDSA process initiated by the 3'-overhang on the genomic S strand **(Figure 6B)**. However, the conversion tracts were slightly wider than those from single-nick-induced SDSA **(Tables S2 and S9)**, probably because the 3' overhang on the S strand exceeds the first nucleotide of the SH mutations by 7 bp **(Figure 4D and Supplemental Sequences)**, which may lead to reduced gene conversion efficiency due to the requirement for relatively extensive end resection. As a result, some PGE products may also stem from the ssDI pathway using the nick on the AS strand, which may contribute to the higher SNP retention frequency on the left side **(Figure 4D)** than is seen in a exclusively SDSA-dependent profile **(Figure 2B)**. Finally, steric hindrance of PAM-in paired-nickases [172-174] may also result in partial nicking in a fraction of cells; and a single nick on the AS strand would have to engage the ssDI pathway. Consistent with all these notions, it is relevant to note that of all the compound lesions and donor ODN configurations analyzed the efficiency of BFP conversion in this PAM-in configuration was the lowest **(Figure 1B)**.

The physical incorporation of ODN donors in the ssDI pathway

The distribution of conversion tracts can be complicated by 3' processing and the co-conversion of proximal SNPs. Besides the directionality of the conversion tracts, a definitive feature that distinguishes the SDSA and ssDI pathways is the physical incorporation of donor ODNs [38, 79]. Thus, the ODNs are physically incorporated to

bridge a genomic lesion in the ssDI pathway, whereas in SDSA, the genomic lesion is repaired by DNA synthesis using the ODN sequence solely as a template **(Figures 3A and 3B)**. To confirm the results of our conversion tract study, we internally labeled a T nucleotide with biotin in S strand ODNs in the center of the SH mutation (TCTCAT, **Supplemental Sequences**), and searched for the physical incorporation of this biotinylated T in the PGE products induced by nCas9, Cas9 and nCas9 paired-nickases **(Figure 7)**. Shortly after transfecting the HPRT-EGFP antisense cells with the designated Cas9 or nCas9, sgRNAs and the biotinylated ODNs, cells with bright BFP expression were isolated by FACS sorting. The genomic DNA from 10^5 BFP positive cells was prepared, digested to completion with XhoI and XbaI restriction enzymes, and pulled down using streptavidin beads [38, 79] **(Figures 7A and 7B)**. The free and non-covalently bond DNA fragments were carefully washed away under denaturing conditions, and the beads with covalently-linked fragments were used as templates in a 40-cycle PCR reaction. An internal primer and a flanking primer were used to specifically detect the ODNs targeted incorporated into the BFP locus (BFP_CF and BFP_QR, **Supplemental Sequences**). Using biotinylated ODNs on the S strand, the covalently-linked genomic fragments were detected only with a single sgRNA on the AS strand and dual sgRNAs in the PAM-in configuration, but not with a single sgRNA on the S strand, a single sgRNA with wildtype Cas9 [79] and dual sgRNAs in the PAM-out configuration **(Figure 7C)**. Collectively, these results confirmed that the ODNs were physically incorporated into the chromosome only when the ssDI model was predicted by the conversion tract profiles **(Figures 2D and 4B)**.

Discussion

Our conversion tract and ODN incorporation studies not only establishes the properties of the individual SDSA and ssDI pathways, but also illustrates the hierarchy of ODN-mediated PGE mediated by double-stranded genomic lesions. The SDSA and ssDI pathways have their respective effective conversion zones, as defined by the region with more than 50% conversion frequency with respect to the genomic lesion initiating the pathways (**Figure 3, bottom**, also see **Figure 2**). Therefore, conversion of knock-in mutations is inefficient outside the conversion zones and this parameter should be considered when designing gene targeting studies. Although both pathways work robustly by themselves **(Figure 1B and Figure S2A)**, it appears that they don't work additively in the presence of DSBs when both pathways are nonetheless in their effective zone [79] **(Figures 2E and 7C)**. Instead, SDSA is preferentially utilized whenever the knock-in mutation is within the effective zone of the 3' overhang complementary to the strand with the ODN donors **(Figures 2E and Figure 4)**. This could be because cells tend to prevent the physical incorporation of exogenous sequences during HDR whenever possible. However, when the SDSA pathway drops out of its effective zone, ssDI is capable of generating the majority of the PGE products **(Figures 4B and 4C)**.

Interestingly, double-stranded lesions generally induced PGE with a higher frequency than single-strand nicks **(Figure 1B)**. The only instance where a complex genomic lesion didn't induce increased PGE was in the case of PAM-in paired-nickases using ODN donors from the AS strand **(Figure 1B)**. As elaborated above, this was also the only instance where extensive (> 7 nt) resection of the 3'-end was required for PGE to occur and we believe that these two observations are connected. If this exception is

illustrative, it may suggest that double-strand lesions in general loosen the local chromosomal architecture more than single-strand nicks and thus permit relevant co-factors (such as resection nucleases) easier access to the 3'-end and ultimately facilitating more vigorous homology searches.

In our nick-induced PGE experiments, no significant strand bias for either the nick or the ODN donors was observed **(Figure 1B)**. These observations were consistent with a previous report using zinc-finger nickases at three independent chromosomal loci [177] but inconsistent with a more recent one [40]. The latter difference might be explained by the fact that the recent report employed ODN donors with a large (17 bp) heterology flanking the genomic lesion [40]. As discussed above, a sizable heterology on both sides of the genomic lesion would strongly favor the use of ssDI as a bidirectional gene conversion pathway. In contrast, in order to be converted by the SDSA pathway, 3' resection of at least 8 to 9 bp would be required and that likely occurs at a much lower frequency. Collectively, these observations suggest that strand-specific transactions such as transcription and DNA replication are unlikely to have a major influence on PGE mediated by nicks. With that said, it is important to note that while no strand-specific biases were noted in any of our experiments, a significant 2- to 4-fold higher frequency of PGE was noted for the BFP reporter in the anti-sense orientation than in the sense orientation **(Figure 1 and Figure S3)**. The reason for this bias is not completely clear, but it is unlikely to be due to conflicting transcription between the EGFP reporter and the endogenous HPRT locus because EGFP expression was very high in these cells. Indeed, we believe the bias was technical (and not biological) in nature since the higher EGFP (and consequently BFP) expression may have simply allowed us to more easily isolate correctly targeted cells.

PAM-in paired-nickases are significantly less effective in producing NHEJ-related mutations than their PAM-out counterparts, which has mainly been attributed to the steric hindrance of the nCas9 proteins in the PAM-in configuration and/or the displacement loops formed the by sgRNAs [172-174]. Our results may provide an additional or alternative explanation **(Figure 5A and 6A)**: PAM-out paired-nickases generate 5'-overhangs that are subject to exonuclease resection [176]. The resection creates a double-strand gap that will normally engage the error-prone NHEJ pathway and produce deletions in between. In contrast, the 3' overhangs produced by PAM-in paired-nickases may remain largely as separated nicks **(Figure 5B and Figure 6B)**, which may be precisely repaired by standard single-strand nick repair without engaging the NHEJ pathway. Importantly, however, the PAM-in paired-nickases can induce PGE at a comparable efficiency to their PAM-out counterparts **(Figure 1B)**, at least when the PAM sequences are placed far enough apart from each other to avoid steric hindrance. Thus, we propose that PAM-in paired-nickases are less efficient in NHEJ-mediated gene disruption because the 3' overhangs rarely form double-strand gaps due to the lack of 3' to 5' resection [176]. Consequently, although PAM-in paired nickases are not often used for PGE, they may actually be advantageous because they may induce similar levels of HDR with less NHEJ mutations compared to their PAM-out counterparts.

Although paired-nickases induced PGE with a higher frequency compared to a single nickase **(Figure 1B)**, we noticed that paired-nickases (in both PAM configurations) occasionally produced imprecise HDR products using ODN donors **(Tables S6-9)**. Most of the imprecise products contained NHEJ-like mutations near the position of the distal nick with respect to the knock-in mutations. Since a single nickase can also generate NHEJ mutations [40, 80], we propose that the imprecise HDR was generated by the repeated nicking of the distal nickase in the PGE products. These mutations were not

found near the position of the proximal nicks because, in our system, the conversion of the knock-in mutations would destroy the binding site of the proximal nickase and prevent further nicking in the PGE products. If this hypothesis is correct, an important guideline of ODN-mediated PGE would be to introduce an additional silent mutation in the PAM sequence of the ODN donors to prevent the further binding of Cas9 or nCas9 in the PGE products.

Our conversion tract data also provides insight into a recent paradox: although SDSA is the more efficient form of meganuclease-induced HDR [126, 178], the majority of the PGE products are actually generated by the DSBR pathway when dsDNA donors with long central heterologies are employed [125, 126, 169, 179]. Our results and other's demonstrate that SDSA is apparently a short-tract gene conversion pathway using chromosomal [180, 181] and ODN donors **(Figures 2A and 2B)**. Also, the Gaussian distribution nature of the associated conversion tracts makes it extremely inefficient in producing gene conversions longer than 100 bp **(Figures 2A and 2B)**. Thus, dsDNA donors with long central heterologies would perforce engage a long-tract gene conversion pathway in the PGE products. Notably, meganuclease-induced HDR using chromosomal donors occasionally generated unidirectional long-tract gene conversion products [180, 181], which were believed to be produced by break-induced replication (BIR) [34-36]. In the case of PGE, however, BIR using exogenous homology donors leads to the formation of chromosome-donor fusions and severely compromises genomic stability [34-36]. In contrast, our previous work demonstrated that the DSBR pathway is a second (and precise) long-tract gene conversion pathway [169], which generates bidirectional conversion tracts with a linear distribution. Collectively, these results indicate that although SDSA is the more efficient HDR pathway for short-tract

gene-conversion, the DSBR pathway is the predominant pathway of PGE using dsDNA donors with long central heterologies.

dsDNA donors can mediate robust PGE in our hand. We can routinely achieve about 20% marker-free PGE using Cas9-2A-GFP with associated FACS enrichment [182] and circular dsDNA donors with long central heterology (data not shown). As a practical guideline, we propose that circular dsDNA should be utilized in the presence of DSBs and/or paired-nicks for converting large knock-in mutations. One of the reasons that these donors work so well is that they may have a longer half-life than ODN donors, which better coincides with the kinetics of the Cas9 expressed in a DNA format. ODN donors should be utilized for converting small modifications. Although ODN donors are optimally used for introducing SNPs, they can also be used to introduce more complex genomic lesions by producing slightly longer conversion tracts. Regardless, what is critical is that the knock-in mutation(s) needs to be placed within the effective conversion zones of the SDSA or ssDI pathways, or it will not be incorporated into the genome. Cas9 expressed in mRNA or protein formats may be used to accommodate the shorter half-life of ODN donors and increase the efficiency of PGE.

In conclusion, our studies have proved the SDSA and ssDI models of ODN-mediated PGE [40] with two pieces of molecular evidence: the directionality of conversion tracts and the physical incorporation of the donor ODNs. In contrast to the DSBR model [169], both pathways produce short conversion tracts with Gaussian distributions. We also extended these observations to compound genomic lesions, and demonstrated that SDSA is preferentially utilized when the knock-in mutation lies within its effective conversion zone, whereas ssDI serves as a backup pathway. The identity, property and hierarchy of these pathways should serve as general guidelines of donor design in any PGE experiment: dsDNA donors are required for introducing large knock-

in mutations, whereas ODNs can conveniently convert small modifications; all the mutations have to be engineered within the effective conversion zones of the specific pathways. These findings also advance our understanding of endogenous HDR-mediated gene conversions, which, in turn, should shed light on potential improvements of PGE.

Figures and Legends

Figure 1. The EGFP > BFP conversion system in the HPRT-EGFP antisense cell line

(A) Schematics of the EGFP > BFP conversion system in the HPRT-EGFP antisense cell line. An EGFP expression cassette was knocked into the HPRT locus of the HCT116 cell line (only the antisense orientation is diagrammed here for the sake of ease of presentation). Genomic lesions could then be introduced near the chromophore (TY residues) of the EGFP sequence using CRISPR/Cas9 cleavase, nickases or dual nickases. HDR repair of the genomic lesions using ODN donors containing the sequence of the SH residues leads to the conversion of EGFP to BFP. The ODN donors also contained synonymous SNPs that could be co-converted with the sequence of the SH residues. Schematic elements: EGFP, green boxes; BFP, blue boxes; HPRT, inverted yellow boxes; sequences of the chromophore residues, CY and SH; the cytomegalovirus promoter, CMV; the polyadenylation sequence, pA; Cas9 variants, red ovals with scissors; genomic lesions, lightning bolts; ODN donors, horizontal red lines; synonymous SNPs, vertical blue hashmarks. (B) The efficiency of BFP conversion in the HPRT-EGFP antisense cell line. The wildtype Cas9 and D10A variant are labeled as Cas9 and nCas9, respectively. The strandedness (S, sense; AS, antisense) of the sgRNA and ODNs (for this experiment without synonymous SNPs) are labeled with respect to the coding sequence of EGFP. Note that the Cas9 D10A variant nicks the strand complementary to that of the sgRNA. All data are shown as the mean ± SEM of three biological replicates. (C) Single representative flow cytometry images of each bar shown in (B).

Figure 1

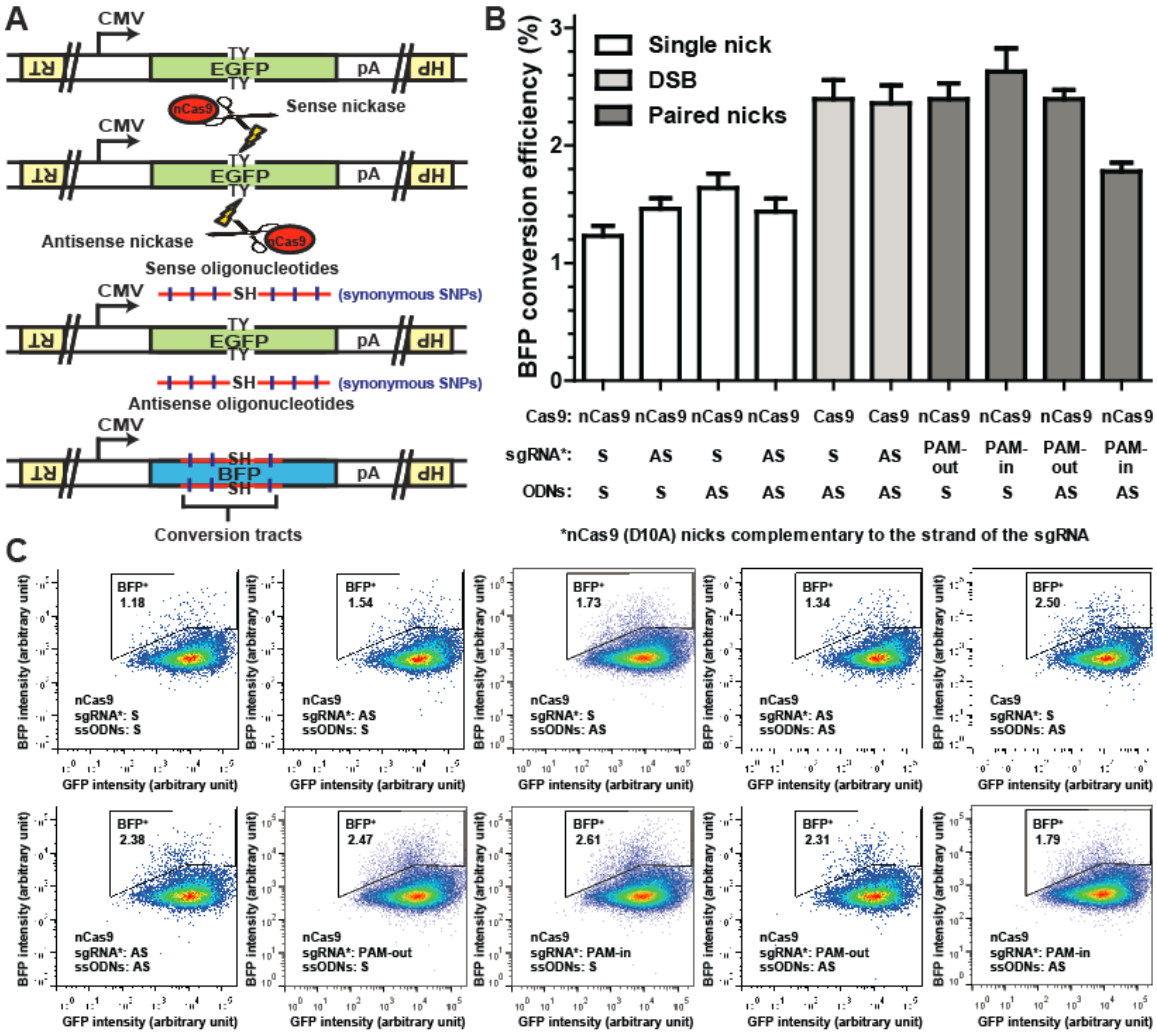

A
CMV
RT
TY
EGFP
TY
pA
dH
nCas9
Sense nickase
Antisense nickase
Sense oligonucleotides
SH
(synonymous SNPs)
Antisense oligonucleotides
BFP
Conversion tracts
B
Single nick
DSB
Paired nicks
BFP conversion efficiency (%)
Cas9: nCas9 nCas9 nCas9 nCas9 Cas9 Cas9 nCas9 nCas9 nCas9 nCas9
sgRNA*: S AS S AS S AS PAM-out PAM-in PAM-out PAM-in
ODNs: S S AS AS AS AS S S AS AS
*nCas9 (D10A) nicks complementary to the strand of the sgRNA
C
BFP intensity (arbitrary unit)
GFP intensity (arbitrary unit)
BFP+ 1.18
nCas9 sgRNA*: S ssODNs: S
BFP+ 1.54
nCas9 sgRNA*: AS ssODNs: S
BFP+ 1.73
nCas9 sgRNA*: S ssODNs: AS
BFP+ 1.34
nCas9 sgRNA*: AS ssODNs: AS
BFP+ 2.50
Cas9 sgRNA*: S ssODNs: AS
BFP+ 2.38
nCas9 sgRNA*: AS ssODNs: AS
BFP+ 2.47
nCas9 sgRNA*: PAM-out ssODNs: S
BFP+ 2.61
nCas9 sgRNA*: PAM-in ssODNs: S
BFP+ 2.31
nCas9 sgRNA*: PAM-out ssODNs: AS
BFP+ 1.79
nCas9 sgRNA*: PAM-in ssODNs: AS

Figure 2. Conversion tracts of single-nick and DSB-induced HDR using ODN donors

(A-E) The conversion tracts of single-nick-induced HDR using ODN donors complementary to (A, B) or the same strand of (C, D) the nick, and DSB-induced HDR (E). The conversion tracts were compiled by overlaying the retention frequency of each SNP in both 6SNP_A and 6SNP_B donors. The positions of the SNPs and predicted genomic lesions are labeled in reference to the center of the chromophore sequence. Schematic elements are colored as follows: genomic DNA, black; genomic lesions, orange; ODNs, red; chromophore sequences, TY and SH; SNPs on the 6SNP_A ODNs, blue; SNPs on the 6SNP_B ODNs, green; homology regions, dashed silver crosses; strandedness of DNA, S and AS. (F) Schematics of the ODN donors with synonymous SNPs. The SNPs are represented as the central nucleotide of the tri-nucleotides in the S orientation, and the distance between the SNPs and the center of the SH sequence (TCTCAT) is labeled. Because the last T in the SH sequence is a wobble nucleotide it could be used as SNP as well and is represented as the ATG at position +3. ODNs with six distributed or clustered SNPs are labeled as 6SNP_A (A) and 6SNP_B (B), respectively. The sequences shown at the bottom of the panel emphasize that the S nick is made at position -1 and the AS nick at position -2. The actual sequences of the ODNs can be found in Supplemental Sequences.

Figure 2A-D

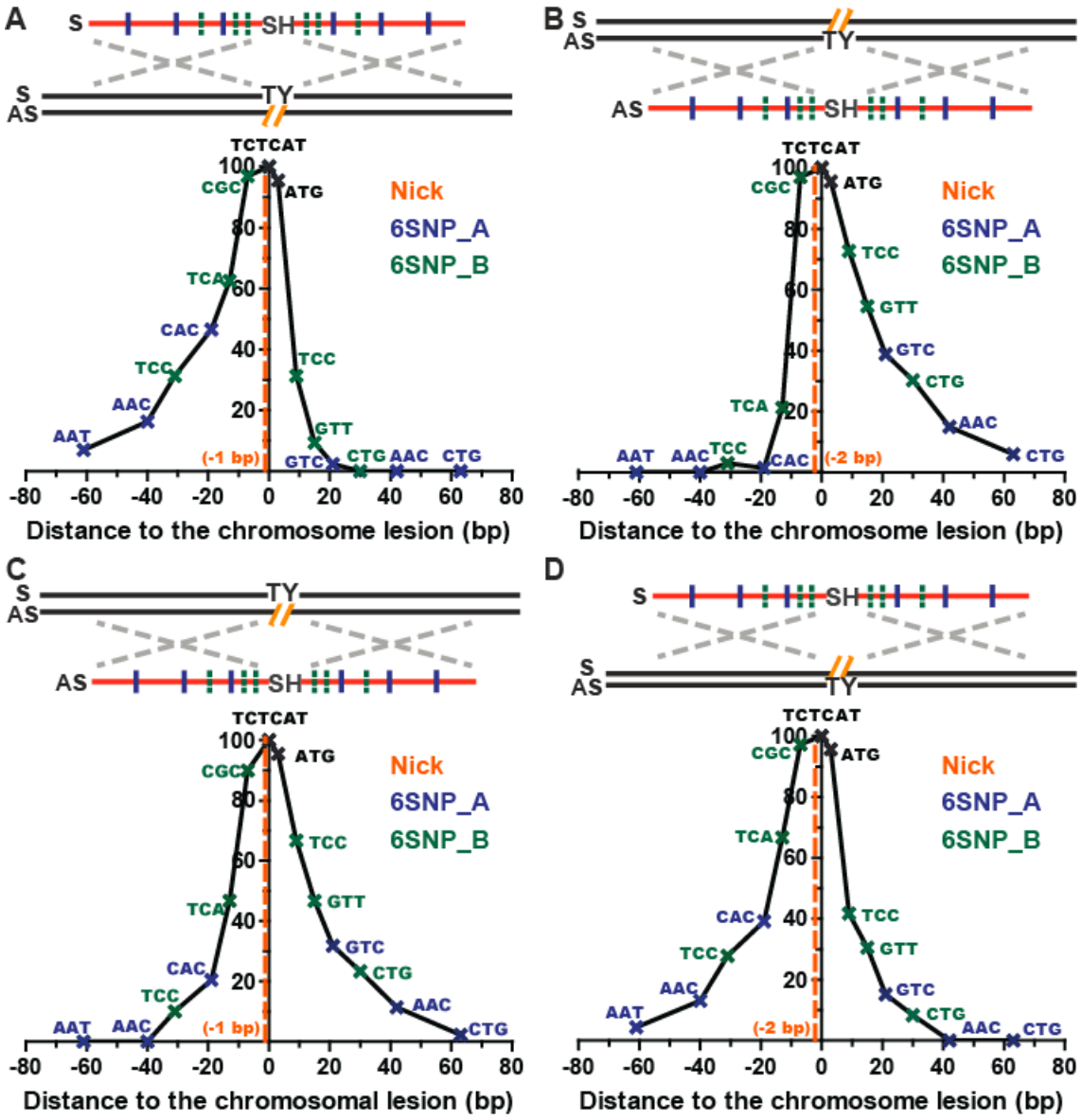

A
S
SH
S
AS
TY
TCTCAT
CGC
ATG
TCC
GTT
GTC
CTG
AAC
CTG
TCA
CAC
TCC
AAC
AAT
(-1 bp)
Nick
6SNP_A
6SNP_B
Distance to the chromosome lesion (bp)
B
S
AS
TY
AS
SH
TCTCAT
CGC
ATG
TCC
GTT
GTC
CTG
AAC
CTG
TCA
AAT
AAC
TCC
CAC
(-2 bp)
Nick
6SNP_A
6SNP_B
Distance to the chromosome lesion (bp)
C
S
AS
TY
AS
SH
TCTCAT
CGC
ATG
TCC
GTT
GTC
CTG
AAC
CTG
TCA
CAC
TCC
AAT
AAC
(-1 bp)
Nick
6SNP_A
6SNP_B
Distance to the chromosomal lesion (bp)
D
S
SH
S
AS
TY
TCTCAT
CGC
ATG
TCC
GTT
GTC
CTG
AAC
CTG
TCA
CAC
TCC
AAC
AAT
(-2 bp)
Nick
6SNP_A
6SNP_B
Distance to the chromosome lesion (bp)

Figure 2E-F

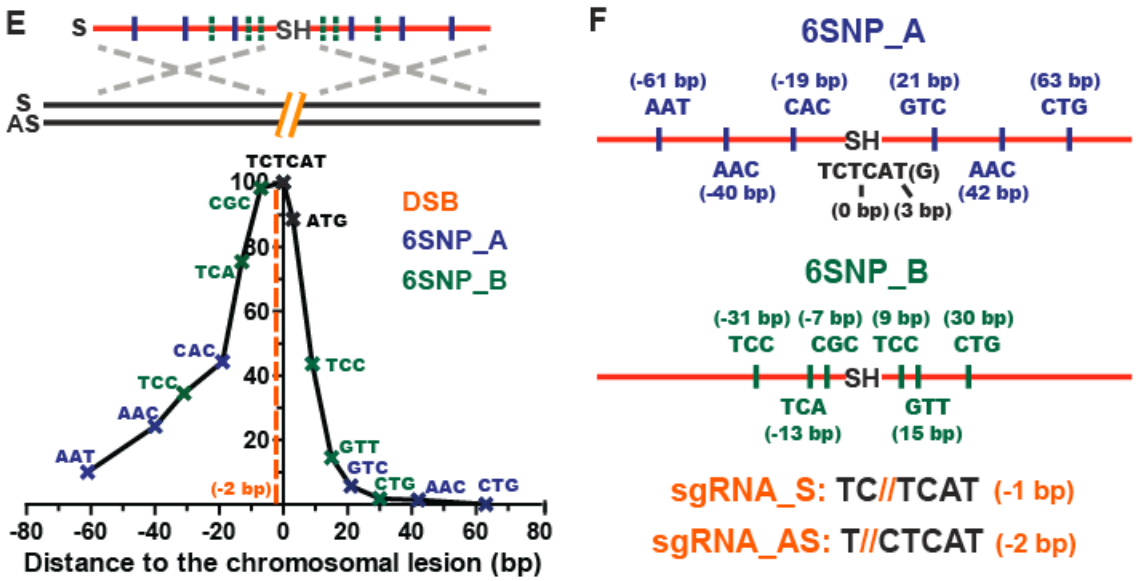

E
S
SH
S
AS
TCTCAT
100
CGC
ATG
80
TCA
60
CAC
TCC
40
AAC
TCC
AAT
20
GTT
GTC
CTG
AAC
CTG
(-2 bp)
-80
-60
-40
-20
0
20
40
60
80
Distance to the chromosomal lesion (bp)
DSB
6SNP_A
6SNP_B
F
6SNP_A
(-61 bp)
AAT
(-19 bp)
CAC
(21 bp)
GTC
(63 bp)
CTG
SH
AAC
(-40 bp)
TCTCAT(G)
(0 bp) (3 bp)
AAC
(42 bp)
6SNP_B
(-31 bp)
TCC
(-7 bp)
CGC
(9 bp)
TCC
(30 bp)
CTG
SH
TCA
(-13 bp)
GTT
(15 bp)
sgRNA_S: TC//TCAT (-1 bp)
sgRNA_AS: T//CTCAT (-2 bp)

Figure 3. Mechanisms of single-nick-induced PGE using ODN donors

(A) Repair of complementary strand nicks via SDSA. (B) Repair of nicks on the same strand via ssDI. Schematic elements are labeled as follows: Top — genomic DNA, black; genomic lesions and DNA ends, hatched orange lines; ODNs, red; knock-in mutations, green; resection nucleases, yellow PAC Man^{TM}; base pairing, purple vertical lines; DNA synthesis, red dashed arrows; processing nucleases, yellow lightning bolts; Bottom — predicted SNP retention curves, red dashed line; predicted position of genomic lesions, orange vertical dashed line; position of the knock-in mutation selected for, green vertical solid line; regions with more than 50% co-conversion frequency, black double-headed arrows, strandedness of DNA, S and AS; orientations of the DNA ends, 5' and 3'.

Figure 3

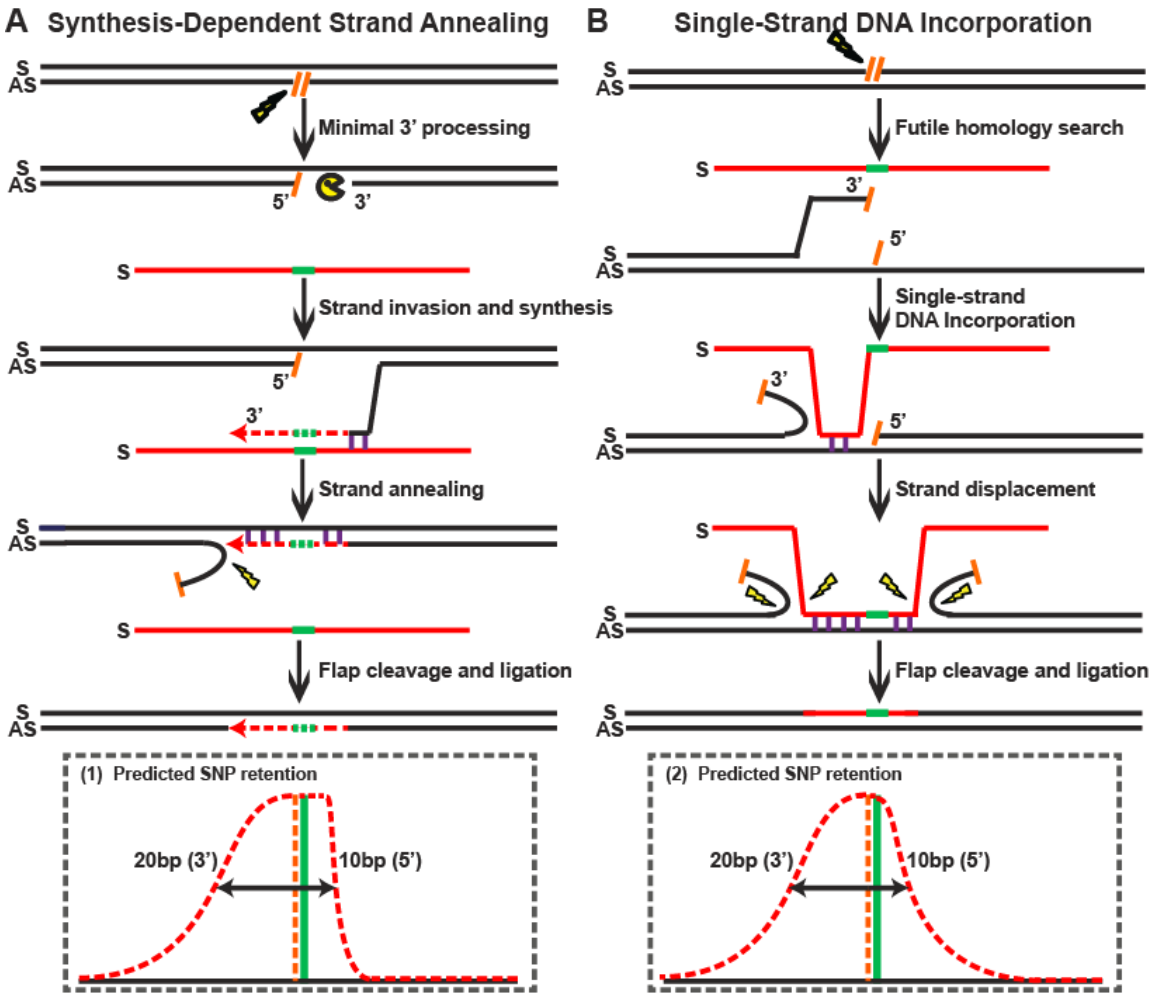

A Synthesis-Dependent Strand Annealing
S
AS
Minimal 3' processing
5'
3'
Strand invasion and synthesis
Strand annealing
Flap cleavage and ligation
(1) Predicted SNP retention
20bp (3')
10bp (5')
B Single-Strand DNA Incorporation
Futile homology search
Single-strand
DNA Incorporation
Strand displacement
Flap cleavage and ligation
(2) Predicted SNP retention
20bp (3')
10bp (5')

Figure 4. The conversion tracts of paired-nicks-induced HDR and the physical incorporation of ODN donors

(A-D) The conversion tracts of paired-nicks-induced HDR using S (A, B) and AS ODNs (C, D). The conversion tracts were compiled by overlaying the retention frequency of each SNP in both 6SNP_A and 6SNP_B donors **(Figure 2F)**. The positions of the SNPs and predicted genomic lesions are labeled in reference to the center of the chromophore sequence. All schematic elements are colored as in the legend of **Figure 2**.

Figure 4

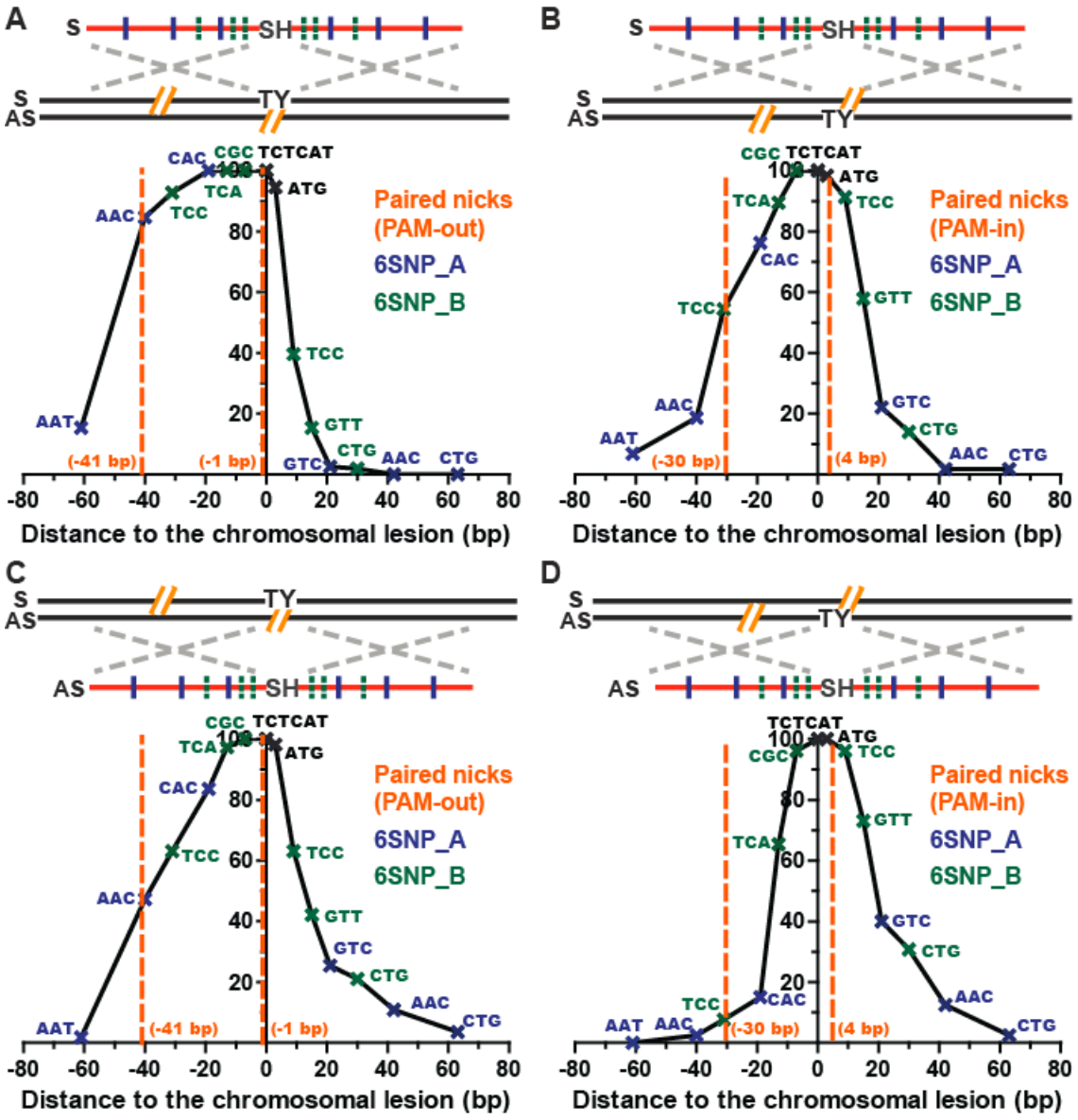

A
S
SH
S
AS
TY
CAC
CGC
TCTCAT
TCA
ATG
AAC
TCC
Paired nicks
(PAM-out)
6SNP_A
6SNP_B
TCC
AAT
GTT
CTG
(-41 bp)
(-1 bp)
GTC
AAC
CTG
100
80
60
40
20
-80 -60 -40 -20 0 20 40 60 80
Distance to the chromosomal lesion (bp)
B
S
SH
S
AS
TY
CGC
TCTCAT
ATG
TCA
TCC
Paired nicks
(PAM-in)
6SNP_A
6SNP_B
CAC
TCC
GTT
AAC
GTC
AAT
CTG
(-30 bp)
(4 bp)
AAC
CTG
C
S
TY
AS
AS
SH
CGC
TCTCAT
TCA
ATG
CAC
Paired nicks
(PAM-out)
6SNP_A
6SNP_B
TCC
TCC
AAC
GTT
GTC
CTG
AAC
AAT
CTG
(-41 bp)
(-1 bp)
D
S
TY
AS
AS
SH
TCTCAT
ATG
CGC
TCC
Paired nicks
(PAM-in)
6SNP_A
6SNP_B
GTT
TCA
GTC
CTG
CAC
AAC
TCC
AAT
AAC
(-30 bp)
(4 bp)
CTG

Figure 5. Mechanisms of paired-nick-induced PGE using the S ODNs

(A) Repair of the PAM-out double nicks via SDSA. (B) Repair of the PAM-in double nicks via ssDI. All schematic elements are colored and defined as in the legend to **Figure 3**.

Figure 5

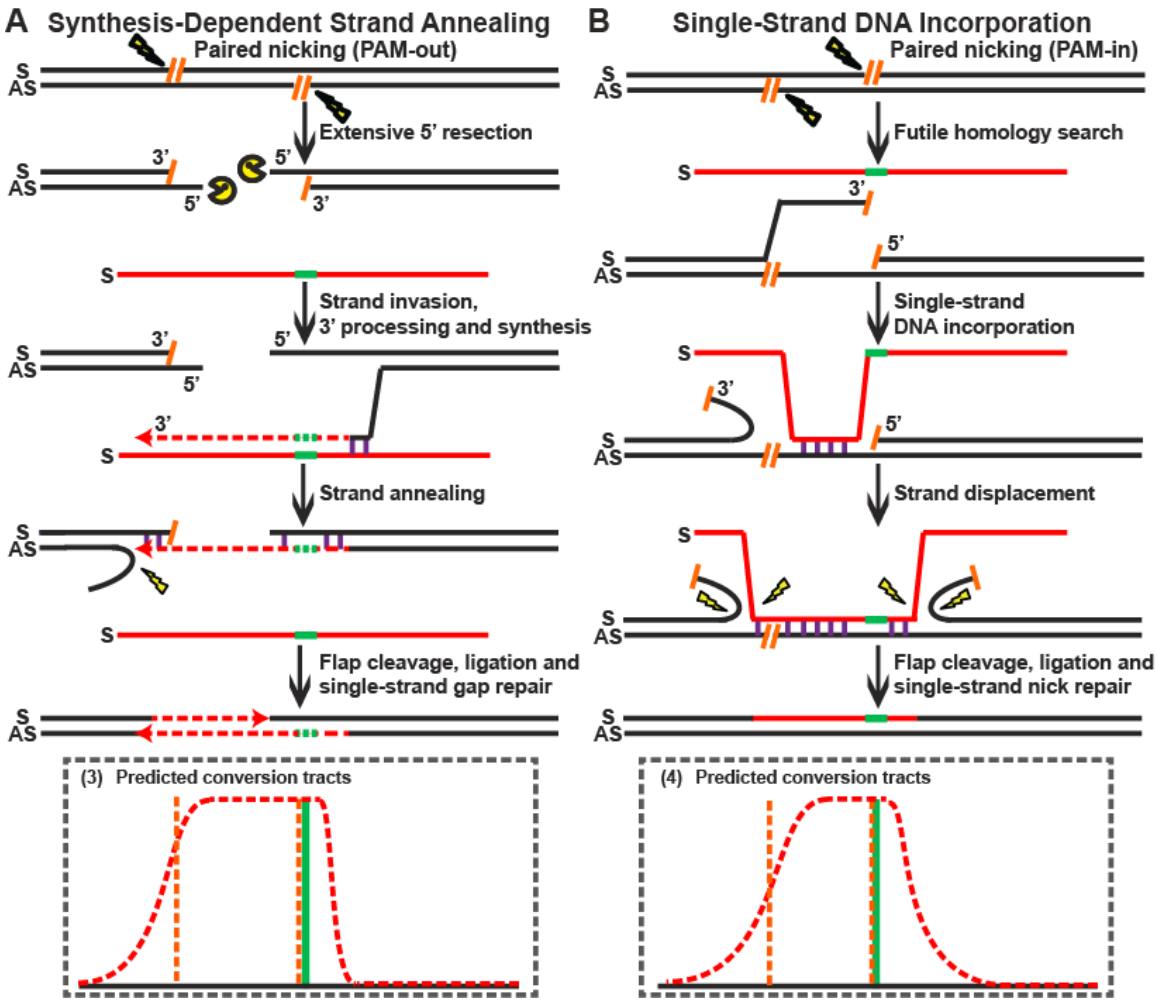

A Synthesis-Dependent Strand Annealing
Paired nicking (PAM-out)
Extensive 5' resection
Strand invasion,
3' processing and synthesis
Strand annealing
Flap cleavage, ligation and
single-strand gap repair
(3) Predicted conversion tracts
B Single-Strand DNA Incorporation
Paired nicking (PAM-in)
Futile homology search
Single-strand
DNA incorporation
Strand displacement
Flap cleavage, ligation and
single-strand nick repair
(4) Predicted conversion tracts
S
AS
3'
5'

Figure 6. Mechanisms of paired-nick-induced PGE using the AS ODNs

(A) Repair of the PAM-out double nicks via ssDI. (B) Repair of the PAM-in double nicks via SDSA. All schematic elements are colored and defined as in the legend to **Figure 3**.

Figure 6

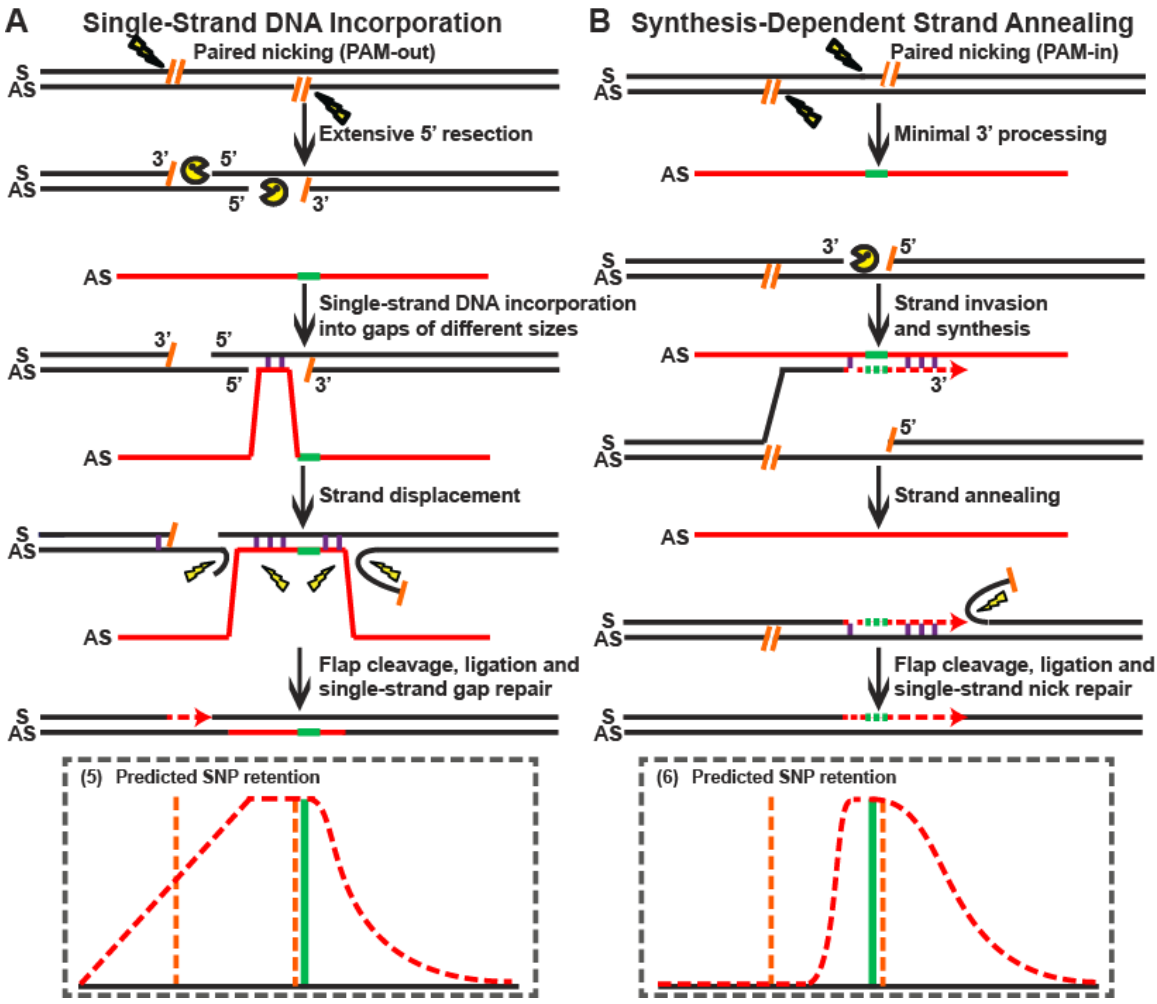

A
Single-Strand DNA Incorporation
Paired nicking (PAM-out)
Extensive 5' resection
Single-strand DNA incorporation into gaps of different sizes
Strand displacement
Flap cleavage, ligation and single-strand gap repair
(5) Predicted SNP retention
B
Synthesis-Dependent Strand Annealing
Paired nicking (PAM-in)
Minimal 3' processing
Strand invasion and synthesis
Strand annealing
Flap cleavage, ligation and single-strand nick repair
(6) Predicted SNP retention
S
AS
3'
5'

Figure 7. The biotin pull-down assay

(A, B) Schematic illustration of the biotin pull-down assay via the SDSA (A) and ssDI (B) pathways. In the case of ssDI (A) the biotinylated ODN is predicted to be incorporated into the target genomic locus whereas in SDSA (B) it should not. The XhoI and XbaI digested genomic fragments covalently linked to biotin can be enriched using streptavidin beads under denaturing conditions. The primers BFP_QF and BFP_QR can specifically amplify these genomic fragments with ODN incorporation but not free ODN donors. Biotin, yellow circle; streptavidin, orange ovals; genomic DNA, black lines; ODN sequence, solid red lines; DNA synthesis, dashed red lines; chromophore sequence, TY and SH; genomic lesions, hatched yellow lines; homology regions, dashed silver crosses; restriction sites, XhoI and XbaI; PCR primers, horizontal arrows. (C) Results of the biotin pull-down assay. After transfecting with biotinylated ODNs (BFP_S90_Biotin, **Supplemental Sequences**), the BFP-positive cells were enriched using FACS sorting. The genomic DNA of the BFP-positive cells was digested with XhoI and XbaI. The fragments covalently linked to biotin were pulled down with streptavidin beads in denaturing conditions, PCR amplified using one internal primer and one flanking primer of the ODN donors (BFP_QF and BFP_QR, **Supplemental Sequences**) and detected on an agarose gel.

Figure 7

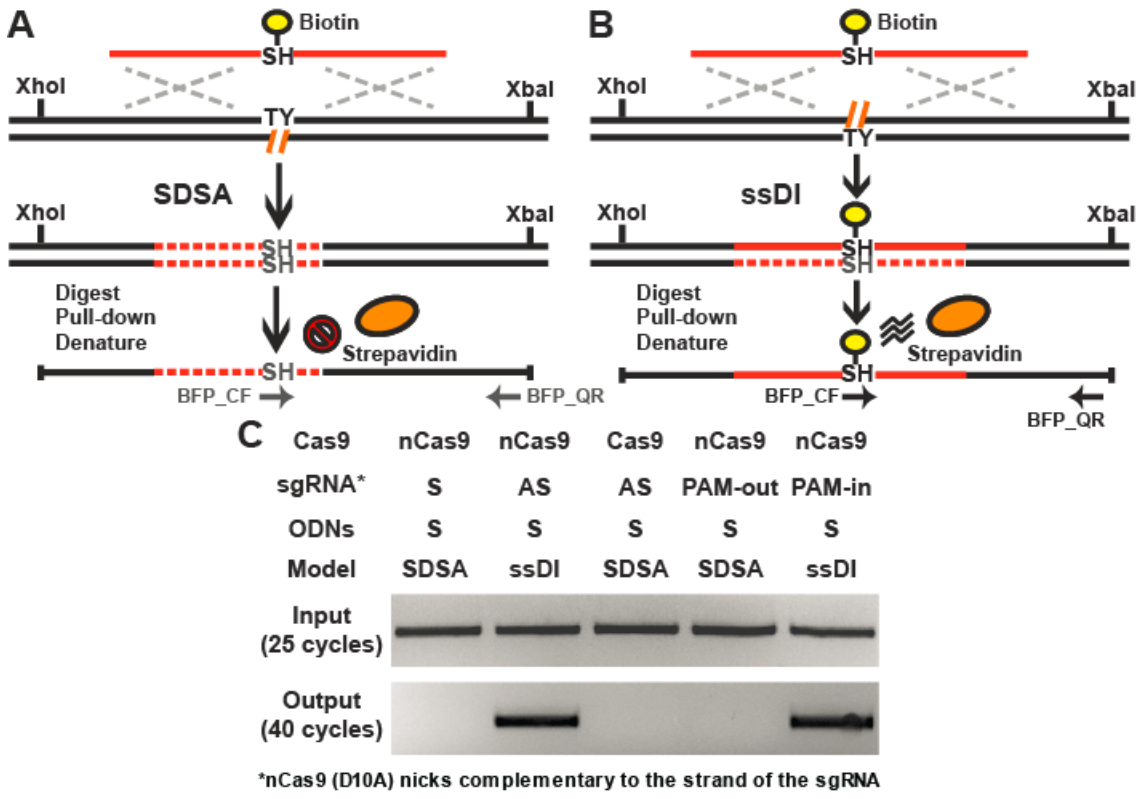

A
Biotin
SH
XhoI
XbaI
TY
SDSA
Digest
Pull-down
Denature
Strepavidin
BFP_CF
BFP_QR
B
ssDI
C
Cas9
nCas9
nCas9
Cas9
nCas9
nCas9
sgRNA*
S
AS
AS
PAM-out
PAM-in
ODNs
S
S
S
S
S
Model
SDSA
ssDI
SDSA
SDSA
ssDI
Input
(25 cycles)
Output
(40 cycles)
*nCas9 (D10A) nicks complementary to the strand of the sgRNA

Figure S1. Schematics of the HDR pathways leading to PGE

(A) Synthesis-dependent strand annealing. (B) Single-strand DNA incorporation. (C) Double-strand break repair (*aka*, the Szostak model). (D) Holliday junction dissolution. Schematic elements are colored as follows: genomic DNA, black; genomic lesions, hatched orange lines with yellow lightning bolts; ODNs, red; knock-in mutations, green; base-pairing, vertical purple lines; DNA synthesis, dashed arrows; strandedness of DNA, S and AS.

Figure S1

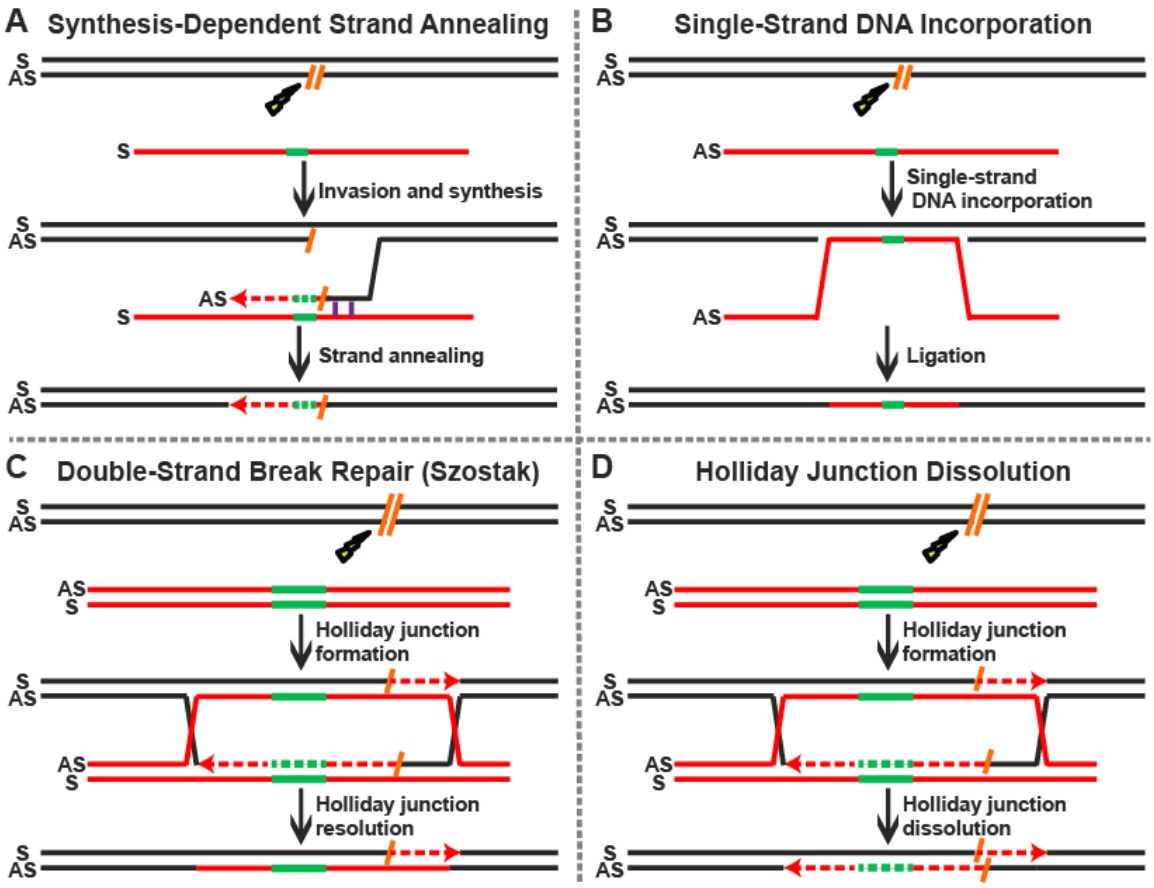

A Synthesis-Dependent Strand Annealing
S
AS
Invasion and synthesis
Strand annealing
B Single-Strand DNA Incorporation
Single-strand
DNA incorporation
Ligation
C Double-Strand Break Repair (Szostak)
Holliday junction
formation
Holliday junction
resolution
D Holliday Junction Dissolution
Holliday junction
formation
Holliday junction
dissolution

Figure S2. Construction of the HPRT-EGFP cell lines

(A) The HPRT-EGFP sense cell line. (B) The HPRT-EGFP antisense cell line. The CMV-EGFP-pA cassette was targeted to replace exon 3 of the HPRT gene in either sense or antisense direction in respect to HPRT using rAAV-mediated gene targeting. Targeted cells were enriched with 6-thioguanine selection and screened using the designated primer pairs (Supplemental sequences). HPRT, hypoxanthine-guanine phosphoribosyltransferase; HA, homology arms; CMV, the cytomegalovirus promoter; pA, the polyadenylation sequence, hairpins, inverted terminal repeats.

Figure S2

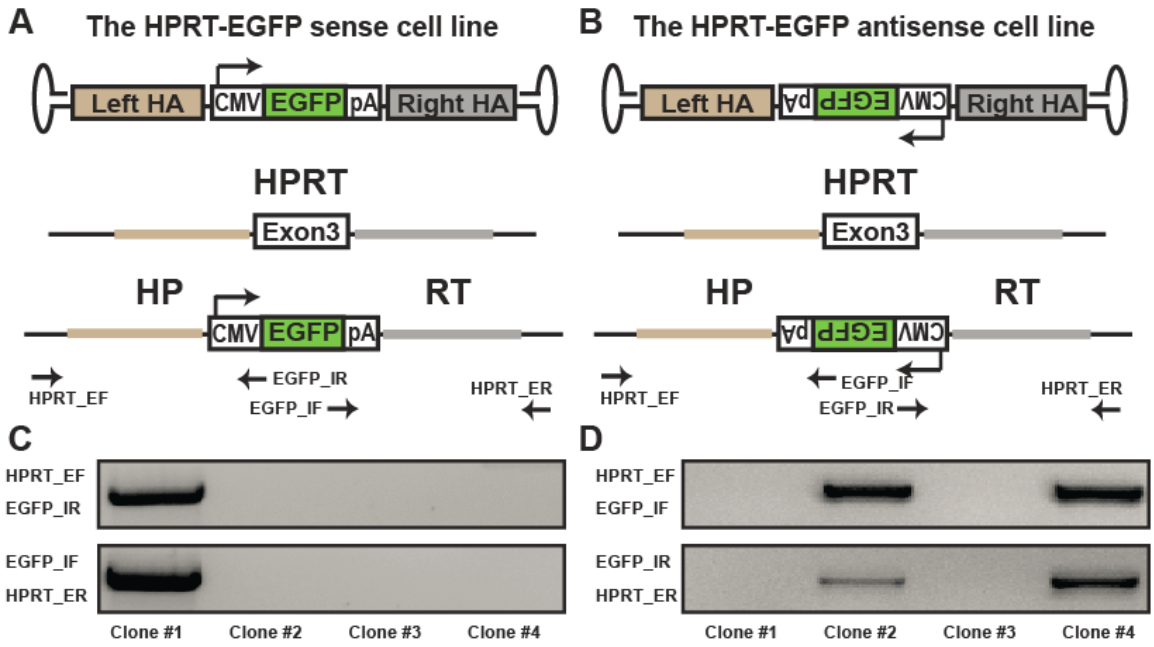
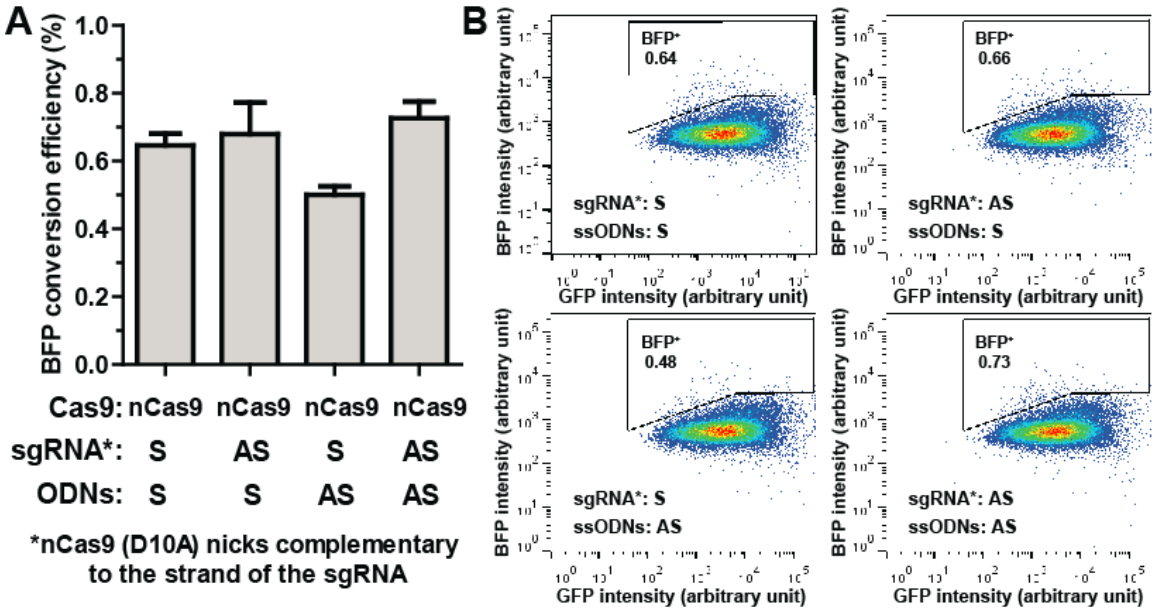

A
The HPRT-EGFP sense cell line
Left HA
CMV
EGFP
pA
Right HA
HPRT
Exon3
HP
RT
HPRT_EF
EGFP_IR
EGFP_IF
HPRT_ER
B
The HPRT-EGFP antisense cell line
Left HA
pA
EGFP
CMV
Right HA
HPRT
Exon3
HP
RT
HPRT_EF
EGFP_IF
EGFP_IR
HPRT_ER
C
HPRT_EF
EGFP_IR
EGFP_IF
HPRT_ER
Clone #1
Clone #2
Clone #3
Clone #4
D
HPRT_EF
EGFP_IF
EGFP_IR
HPRT_ER
Clone #1
Clone #2
Clone #3
Clone #4

Figure S3. The efficiency of BFP conversion in the HPRT-EGFP sense cell line

(A) The efficiency of BFP conversion. The wildtype Cas9 and D10A variant are labeled as Cas9 and nCas9, respectively. The strandedness of the sgRNA and ODNs (without synonymous SNPs) are labeled with respect to the coding sequence of EGFP. Note that the Cas9 D10A variant nicks the strand complementary to that of the sgRNA. All data are shown as the mean ± SEM of three biological replicates. (B) Single representative flow cytometry images of each bar shown in (A). The medians of three biological replicates are shown.

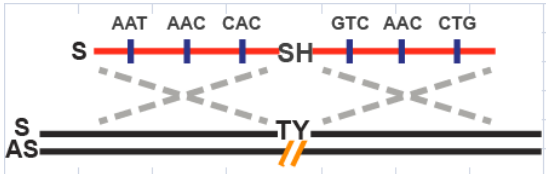
Figure S3


A
BFP conversion efficiency (%)
1.0
0.8
0.6
0.4
0.2
0.0
Cas9: nCas9 nCas9 nCas9 nCas9
sgRNA*: S AS S AS
ODNs: S S AS AS
*nCas9 (D10A) nicks complementary to the strand of the sgRNA
B
BFP intensity (arbitrary unit)
GFP intensity (arbitrary unit)
BFP+ 0.64
sgRNA*: S
ssODNs: S
BFP+ 0.66
sgRNA*: AS
ssODNs: S
BFP+ 0.48
sgRNA*: S
ssODNs: AS
BFP+ 0.73
sgRNA*: AS
ssODNs: AS

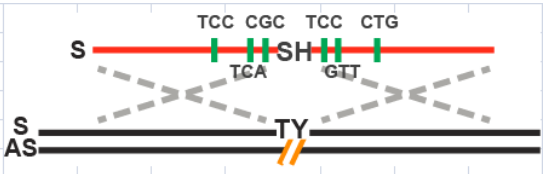
Tables S1-9. Individual conversion tracts of single nick- (S1-4), DSB- (S5) and paired-nick- (S6-9) induced PGE using ODN donors

Top — Schematic configurations of PGE. Genomic DNA, black; genomic lesions, hatched orange lines; chromophore sequences, TY and SH; ODNs, red; SNPs on the 6SNP_A ODNs, blue; SNPs on the 6SNP_B ODNs, green; homology regions, dashed silver crosses; strandedness of DNA, S and AS. Bottom — SNP retention of individual conversion tracts. SNP, center of the trinucleotides; TCTCAT, the SH sequence; retained SNPs, +; lost SNPs, -; retained SNPs on the 6SNP_A ODNs, blue; retained SNPs on the 6SNP_B ODNs, green; number of inserted nucleotides near the designated SNPs, Ins #; number of deleted nucleotides near the designated SNPs, del #.

Table S1. Conversion tracts of nick-induced HDR using S sgRNA* and S ODNs

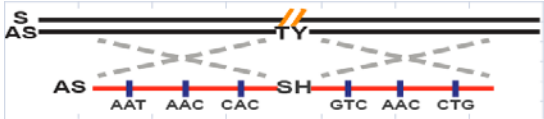


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
3	7	20	43	1	0	0
6.98%	16.28%	46.51%	100.00%	2.33%	0.00%	0.00%

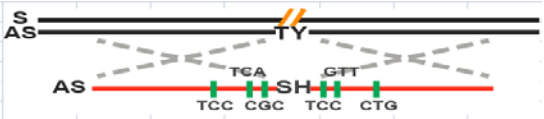


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	-	-	+	+	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	+	+	-
-	+	+	+	-	-	-
-	-	+	+	+	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	+	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	+	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	-	-
-	-	+	+	-	-	-
+	+	+	+	+	-	-
-	+	+	+	+	+	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
10	20	31	32	10	3	0
31.25%	62.50%	96.88%	100.00%	31.25%	9.38%	0.00%

Table S2.Conversion tracts of nick-induced HDR using AS sgRNA*and AS ODNs

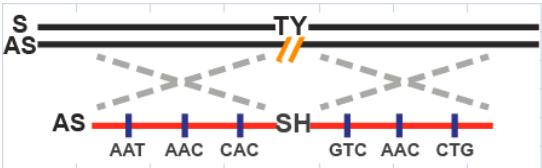


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	+
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	+
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
0	0	1	67	26	10	4
0.00%	0.00%	1.49%	100.00%	38.81%	14.93%	5.97%

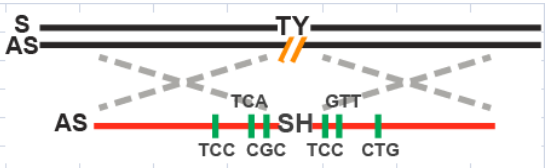


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	+	+	+	-	-	-
-	-	+	+	+	+	-
-	-	+	+	+	+	-
-	-	-	+	-	-	-
-	-	+	+	+	+	-
-	-	+	+	+	+	+
-	-	+	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	+	+	+
-	-	+	+	+	+	+
-	-	+	+	+	-	-
-	-	+	+	+	-	-
-	+	+	+	+	+	-
-	-	+	+	+	+	+
-	+	+	+	+	+	+
-	-	+	+	+	-	-
-	-	+	+	+	+	+
-	-	+	+	+	+	+
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	+	+
-	+	+	+	+	-	-
-	-	+	+	+	-	-
+	+	+	+	+	+	-
-	-	+	+	-	-	-
-	+	+	+	+	+	+
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	+	+
-	-	+	+	+	+	-
-	-	+	+	+	+	-
-	-	+	+	+	+	-
-	-	+	+	+	+	+
-	-	+	+	-	-	-
1	7	32	33	24	18	10
3.03%	21.21%	96.97%	100.00%	72.73%	54.55%	30.30%

Table S3. Conversion tracts of nick-induced HDR using S sgRNA* and AS ODNs

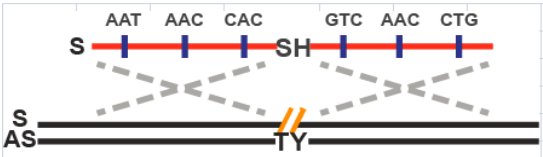


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	+	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	+	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	+
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	+	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
0	0	9	44	14	5	1
0.00%	0.00%	20.45%	100.00%	31.82%	11.36%	2.27%

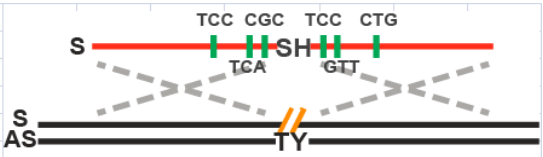


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	+	+	+	-	-	-
-	+	-	+	+	+	+
-	-	-	+	+	+	+
-	-	+	+	+	-	-
-	+	+	+	+	+	-
-	-	+	+	-	-	-
-	-	+	+	+	+	+
-	+	+	+	+	+	-
-	-	+	+	-	-	-
-	-	-	+	+	+	-
+	+	+	+	+	+	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	+	+	-
-	+	+	+	+	-	-
-	-	+	+	+	+	+
-	+	+	+	+	+	+
-	-	+	+	-	-	-
-	+	+	+	+	+	-
-	-	+	+	+	+	+
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	+	+	+
-	+	+	+	+	-	-
-	-	+	+	+	-	-
-	-	+	+	+	-	-
-	+	+	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	+	-
3	14	27	30	20	14	7
10.00%	46.67%	90.00%	100.00%	66.67%	46.67%	23.33%

Table S4. Conversion tracts of nick-induced HDR using AS sgRNA* and S ODNs

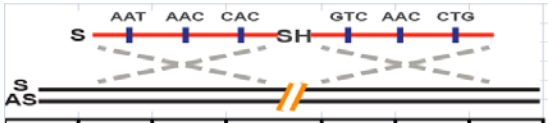


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	-	+	+	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	+	-	-
-	+	+	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	+	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
0	0	0	+	0	0	0
2	6	18	46	7	0	0
4.35%	13.04%	39.13%	100.00%	15.22%	0.00%	0.00%

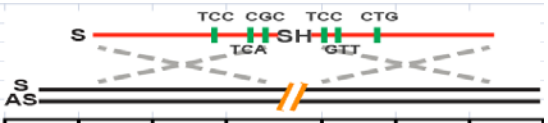


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	+	+	-
-	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	+	+	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	+	+
+	+	+	+	-	-	-
-	-	+	+	+	-	-
+	+	+	+	-	-	-
-	+	+	+	+	+	+
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	+	+
-	-	+	+	+	+	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	+	+	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	+	+	-
-	+	+	+	+	+	-
-	-	+	+	+	+	-
10	24	35	36	15	11	3
27.78%	66.67%	97.22%	100.00%	41.67%	30.56%	8.33%

Table S5. Conversion tracts of DSB-induced HDR using AS sgRNA and S ODNs

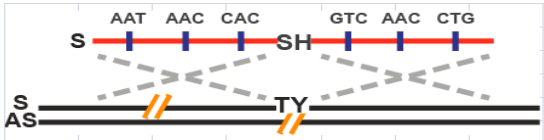


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	-	+	+	+	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
+	+	+	+	+	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
+	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
7	17	31	70	4	1	0
10.00%	24.29%	44.29%	100.00%	5.71%	1.43%	0.00%

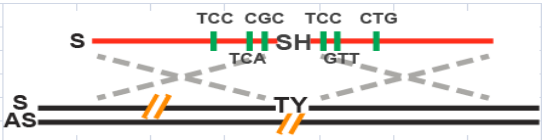


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	+	+	+	-	-	-
-	+	+	+	+	+	-
+	+	+	+	+	+	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	-	-
-	+	+	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	+	+	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	+	-	-
-	+	+	+	-	-	-
+	+	+	+	+	+	-
-	+	+	+	+	+	-
+	+	+	+	-	-	-
-	-	+	+	+	+	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	+	-
-	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	+	+	-
-	-	+	+	-	-	-
-	-	+	+	+	+	+
+	+	+	+	-	-	-
-	-	+	+	+	-	-
-	+	+	+	+	-	-
19	41	54	55	24	11	1
34.55%	74.55%	98.18%	100.00%	43.64%	20.00%	1.82%

Table S6. Conversion tracts of PAM-out paired-nick-induced HDR using S ODNs

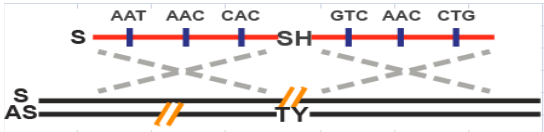


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
6	33	39	39	1	0	0
15.38%	84.62%	100.00%	100.00%	2.56%	0.00%	0.00%

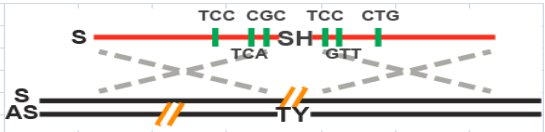


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
Del11	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	+	+
+	+	+	+	+	+	+
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	+	-	-
-	+	+	+	+	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	+	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
-	+	+	+	+	+	-
+	+	+	+	-	-	-
51	55	55	55	23	11	2
92.73%	100.00%	100.00%	100.00%	41.82%	20.00%	3.64%

Table S7. Conversion tracts of PAM-in paired-nick-induced HDR using S ODNs

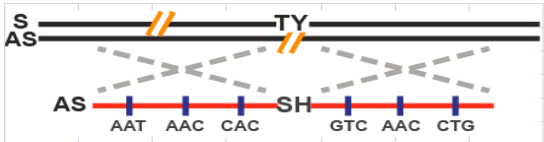


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	-	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	+	-	-
-	-	+	+	+	-	-
-	-	Del34	+	-	-	-
-	-	-	+	+	-	-
-	-	+	+	-	-	-
-	-	-	+	+	-	-
+	+	+	+	-	-	-
-	-	-	+	+	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	Ins23	+	-	-	-
-	+	Del28	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
+	-	+	+	-	-	-
-	-	-	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	+
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	-	-
-	-	+	+	-	-	-
-	-	Ins81	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	+	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	-	-	-
+	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	Del30	-	-	-
-	-	+	+	-	-	-
-	-	-	+	+	+	-
-	-	+	+	-	-	-
-	Del3/Ins5	-	+	-	-	-
4	11	45	59	13	1	1
6.78%	18.64%	76.27%	100.00%	22.03%	1.69%	1.69%

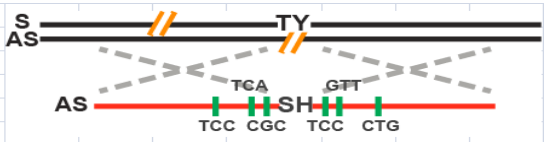


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	-	+	+	+	-	-
+	+	+	+	+	-	-
+	+	+	+	+	+	+
-	+	+	+	+	+	-
+	+	+	+	+	-	-
+	+	+	+	+	+	+
+	+	+	+	+	+	-
-	+	+	+	+	-	-
+	+	+	+	+	-	-
-	+	+	+	+	-	-
+	+	+	+	+	-	-
-	-	+	+	-	+	+
+	+	+	+	+	+	-
-	+	+	+	+	-	-
-	+	+	+	+	+	-
+	+	+	+	+	+	-
-	+	+	+	+	-	-
+	+	+	+	+	+	-
-	+	+	+	+	+	+
+	+	+	+	+	-	-
-	+	+	+	+	-	-
+	+	+	+	+	-	-
-	+	+	+	+	+	-
+	+	+	+	+	+	-
+	+	+	+	+	+	-
-	-	+	+	-	-	-
Del26/Ins25	+	+	+	+	-	-
+	+	+	+	+	-	-
+	+	+	+	+	+	-
+	+	+	+	+	+	-
-	-	+	+	+	+	-
-	+	+	+	+	+	-
-	+	+	+	+	+	-
+	+	+	+	+	+	-
+	+	+	+	+	+	+
-	+	+	+	+	+	-
-	+	+	+	+	-	-
+	+	+	+	-	-	-
-	+	+	+	+	-	-
+	+	+	+	+	+	-
+	+	+	+	+	+	-
-	+	+	+	+	+	+
+	+	+	+	+	-	-
-	-	+	+	+	-	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
-	+	+	+	+	+	+
-	-	+	+	+	+	-
+	+	+	+	+	-	-
+	+	+	+	+	-	-
-	+	+	+	+	+	+
+	+	+	+	+	+	-
+	+	+	+	+	-	-
-	+	+	+	-	-	-
-	+	+	+	+	+	-
+	+	+	+	+	+	-
31	51	57	57	52	33	8
54.39%	89.47%	100.00%	100.00%	91.23%	57.89%	14.04%

Table S8.Conversion tracts of PAM-out paired-nick-induced HDR using AS ODN

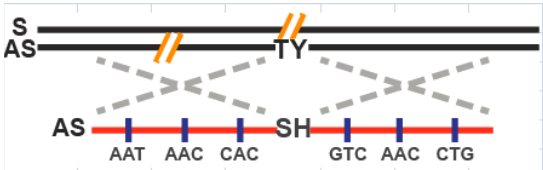


AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	+	+	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	+	+	-
-	+	+	+	+	-	-
-	-	+	+	+	-	-
-	+	+	+	+	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	+	+	+
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	+	+	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	+	+	+
-	-	+	+	+	+	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	+	+	-	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
-	+	+	+	+	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
+	+	-	+	+	-	-
-	-	+	+	-	-	-
-	+	+	+	-	-	-
1	26	46	55	14	6	2
1.82%	47.27%	83.64%	100.00%	25.45%	10.91%	3.64%

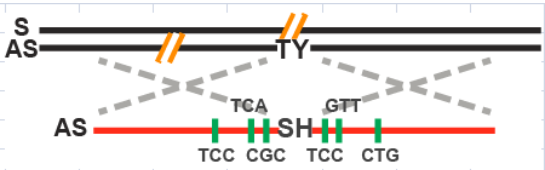


TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
+	+	+	+	-	-	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
-	+	+	+	-	-	-
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+	+	+	+	+	+	+
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	+	-	-
+	+	+	+	+	-	-
+	+	+	+	+	+	-
-	-	+	+	+	+	+
+	+	+	+	+	+	-
-	+	+	+	+	-	-
+	+	+	+	+	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
-	+	+	+	+	-	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	+	+	-
-	+	+	+	-	-	-
-	+	+	+	+	+	+
+	+	+	+	+	-	-
-	+	+	+	+	+	+
+	+	+	+	+	+	-
+	+	+	+	+	+	-
-	+	+	+	-	-	-
+	+	+	+	+	+	-
+	+	+	+	-	-	-
+	+	+	+	-	-	-
-	+	+	+	-	-	-
+	+	+	+	+	-	+
+	+	+	+	+	+	+
+	+	+	+	+	-	-
+	+	+	+	+	+	-
24	37	38	38	24	16	8
63.16%	97.37%	100.00%	100.00%	63.16%	42.11%	21.05%

Table S9. Conversion tracts of PAM-in paired-nick-induced HDR using AS ODNs



AAT	AAC	CAC	TCTCAT	GTC	AAC	CTG
-	-	+	+	+	-	-
-	-	-	+	+	-	-
-	-	Ins29	+	-	-	-
-	-	-	+	-	-	-
-	-	+	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	-	+	+	-	-
-	+	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	Del12	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	+
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	+	-
-	-	+	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	+	+	+	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	+	-
-	-	-	+	+	+	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
-	-	-	Ins29	-	-	-
-	-	-	+	+	-	-
-	-	-	+	-	-	-
-	-	-	+	+	-	-
0	1	6	40	16	5	1
0.00%	2.50%	15.00%	100.00%	40.00%	12.50%	2.50%



TCC	TCA	CGC	TCTCAT	TCC	GTT	CTG
-	+	+	+	+	+	-
-	-	-	+	+	+	+
-	-	+	+	+	+	-
-	+	+	+	+	+	+
-	+	+	+	+	-	-
-	+	+	+	+	+	-
+	+	+	+	+	+	+
-	+	+	+	+	+	+
-	-	+	+	+	+	-
-	-	+	+	+	-	-
-	+	+	+	+	+	-
-	-	+	+	+	+	-
-	+	+	+	+	+	+
-	+	+	+	+	+	-
-	+	+	+	+	-	-
-	+	+	+	+	+	+
-	+	+	+	+	+	-
-	-	+	+	+	-	-
-	+	+	+	+	+	+
-	+	+	+	+	+	-
-	+	+	+	+	-	-
-	-	+	+	-	-	-
+	+	+	+	+	+	-
-	-	+	+	+	+	-
-	+	+	+	+	+	+
-	-	+	+	+	-	-
2	17	25	26	25	19	8
7.69%	65.38%	96.15%	100.00%	96.15%	73.08%	30.77%

Supplemental Sequences

The EGFP reporter cell lines

CMV-EGFP-pA:

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC

CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC

GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC

ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA

GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC

CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA

CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGC

GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT

GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCC

GCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA

GAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGACTCAGATCTCGAG

CTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGCCG

GTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC

CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG

GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCG

GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCA

GTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA

TGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA

CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA

GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG

TACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCAT

CAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC

GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA

ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGA

TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC

GAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTT

GTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACAT

AAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAAT

AAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTG

TGGTTTGTCCAAACTCATCAATGTATCTTAAG

PCR primers for the HPRT homology arms. Only annealing sequences of the primer tails are shown:

HPRT_left_forward (HPRT_LF): CATGTTTGGTACTTGTTCAGC

HPRT_left_reverse (HPRT_LR): TTAGCCAGGCATGGTAGC

HPRT_right_forward (HPRT_RF): ACTAGTCACCTTGGAGGATAT

HPRT_right_reverse (HPRT_RR): CAAAGCATTTCTACCACTCAG

Screening primers:

HPRT_external_forward (HPRT_EF): AGTATCAGTTGTGGTATAGTGG

EGFP_internal_reverse (EGFP_IR): GGTGGTGCAGATGAACTT

EGFP_internal_forward (EGFP_IF): CGACAACCACTACCTGAG

HPRT_external_reverse (HPRT_ER): ATCAGTTGAGGAGTTCAGC

The EGFP-BFP conversion

sgRNA sequences:

GFP_S: CTCGTGACCACCCTGACCTA

GFP_AS: GCACTGCACGCCGTAGGTCA

GFP_PAMout_S: CTCGTGACCACCCTGACCTA

GFP_PAMout_AS: CCAGGGCACGGGCAGCTTGC

GFP_PAMin_S: CCGGCAAGCTGCCCGTGCCC

GFP_PAMin_AS: GCTGAAGCACTGCACGCCGT

ODN donors:

BFP_S160:

ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG

GCCCACCCTCGTGACCACCCTG**TCTCAT**GGCGTGCAGTGCTTCAGCCGCTACCCC

GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTA

BFP_AS160:

TAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGT

AGCGGCTGAAGCACTGCACGCC**ATGAGA**CAGGGTGGTCACGAGGGTGGGCCAGG

GCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGT

BFP_S160_6SNP_A (S_6SNP_A):

ACGGCAAGCTGACCCTGAA**A**TTCATCTGCACCACCGGCAA**A**CTGCCCGTGCCCTG

GCCCAC**A**CTCGTGACCACCCTG**TCTCAT**GGCGTGCAGTGCTTCAG**T**CGCTACCCC

GACCACATGAA**A**CAGCACGACTTCTTCAAGTC**T**GCCATGCCCGAAGGCTA

BFP_S160_6SNP_B (S_6SNP_B):

ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT**C**CCCTG

GCCCACCCTCGT**C**ACCAC**G**CTG**TCTCAT**GGCGT**C**CAGTG**T**TTCAGCCGCTACCC**T**

GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTA

BFP_AS160_6SNP_A (AS_6SNP_A):

TAGCCTTCGGGCATGGCAGACTTGAAGAAGTCGTGCTGTTTCATGTGGTCGGGGT

AGCGACTGAAGCACTGCACGCCATGAGACAGGGTGGTCACGAGTGTGGGCCAGG

GCACGGGCAGTTTGCCGGTGGTGCAGATGAATTTCAGGGTCAGCTTGCCGT

BFP_AS160_6SNP_B (AS_6SNP_B):

TAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCAGGGT

AGCGGCTGAAACACTGGACGCCATGAGACAGCGTGGTGACGAGGGTGGGCCAGG

GGACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGT

BFP_S90_Biotin:

CACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGTC/iBiodT/C

ATGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAA

PCR primers:

BFP_external_forward (BFP_EF): CACAAGTTCAGCGTGTCC

BFP_external_reverse (BFP_ER): GGTGCTCAGGTAGTGGTT

BFP_confirmation_forward (BFP_CF): CCACCCTGTCTCATGGC

BFP_quantification_reverse (BFP_QR): TGTGGCTGTTGTAGTTGTA

Materials and Methods

Nucleotide sequences

All sgRNA targets, ODN donors, primers and the relevant plasmid sequences can be found in **Supplementary Sequences**.

Cell culture

The human HCT116 cell line was cultured in McCoy's 5A medium supplemented with 10% FBS, 1% L-glutamine, 1% penicillin/streptomycin with 5% CO_2 at 37 ^{o}C.

The HPRT-EGFP cell lines

The rAAV EGFP knock-in vectors were constructed using an unpublished method (Kan *et al*., manuscript in preparation). Basically, the CMV-EGFP-pA cassette was amplified from EGFP-N2 (Clontech), and the homology arms flanking HPRT exon 3 were amplified using designated primers **(Supplementary Sequences)**. The CMV-EGFP-pA cassette and homology arms are ligated into a modified version of the pAAV-MCS vector in both sense and antisense orientations with respect to HPRT. rAAV packaging and infections were performed as described [183]. The infected HCT116 cells were seeded into 10 cm tissue culture dishes and selected with 5 μg/mL 6-TG for 14 days. Individual colonies were initially scanned for EGFP expression under a fluorescence microscope, and subsequently screened by PCR using the indicated primers **(Figure S2A, S2B)**. One of the targeted clones with EGFP in either sense or antisense directions in the HPRT locus were flow sorted for high EGFP expression using a FACSAria II cell sorter (BD Biosciences) and used for subsequent studies.

The EGFP > BFP conversion

The EGFP reporter cells were seeded at ~50% confluency. The next day, 5 x 10^5 cells were transfected with the Cas9 or nCas9 expression plasmid (10 μg; #41815 and #41816, Addgene), the MLM3636 plasmid containing designated sgRNA expression cassette (10 μg; #43860, Addgene) and the relevant ODNs (10 μg) using a Neon Transfection System (Invitrogen). For paired-nickases, cells were transfected with 7.5 μg of each sgRNA plasmid, the Cas9 expression plasmid and ODNs. All transfections were performed using 100 μL tips under elevated conditions compared to the manufacturer's protocol: 1530V, a 10 msec pulse width, and 3 pulses. For the PGE efficiency experiment, cells were transfected using ODNs without flanking SNPs **(Supplementary Sequences)**, and the percentage of BFP positive cells were quantitated 2 days after transfection using a LSRFortessa cell analyzer (BD Biosciences). For the conversion tracts experiment, cells were transfected using ODNs containing flanking SNPs. Cells from 5 individual transfections were combined, cultured for 2 days and sorted for BFP expression using a FACSAria II cell sorter (BD Biosciences).

Conversion tracts analysis

The BFP positive cells were single-cell subcloned into 96-well-plates at a concentration of 3 to 10 cells per well. 14 days later, single colonies were trypsinized, transferred into new 96-well plates and grown to confluency. Genomic DNA was prepared using DirectPCR Lysis Reagent (Viagene). The BFP fragments containing all the potential SNPs were amplified and confirmed by PCR using primers BFP_EF X BFP_ER and BFP_CF X BFP_ER, respectively. The PCR products were purified using

a Qiaquick PCR purification kit. The retention of vector-borne SNPs was analyzed by Sanger sequencing.

The biotin incorporation assay

The biotin incorporation assay was performed as described [38] with minor modifications. The EGFP reporter cells were transfected with internal biotin-labeled ODNs (BFP_S90_Biotin) and flow sorted as described above. The genomic DNA from 1 x 10^5 BFP-positive cells was prepared using a Puregene Cell kit (Gentra) with a 60 min centrifugation step after isopropanol precipitation. The genomic DNA was digested to completion with XhoI and XbaI restriction enzymes (New England Biolabs). The biotinylated DNA fragments were isolated using a Dynabeads Kilobase BINDER kit (Life Technologies) according to manufacturer's protocol, except that all reagents were supplemented with 0.1% BSA, and two additional washes using 0.1 M NaOH and 0.05 M NaCl at room temperature and one more rinse with 95 ^{o}C water were performed to remove the non-covalently linked genomic fragments. The biotinylated genomic fragments were detected with a 40-cycle PCR using Phusion Hot Start II High-Fidelity DNA polymerase (Thermo Scientific). The genomic DNA preparation prior to Dynabeads purification was used as control in a 25-cycle reaction.

CHAPTER IV: FINAL DISCUSSIONS

Conclusions

Since CRISPR/Cas9 was repurposed for genome engineering in 2013 [56-60], it has quickly swept through the laboratories around the world and caused a major upheaval in biomedical research [94-99]. For the first time, scientists are able to engineer, in principle, any part of the human genome with extreme ease and precision. If we compare the human genome to an encyclopedia with 24 volumes (unique chromosomes), ~20,000 chapters (genes) and 3,200,000,000 letters (base pairs), the CRISPR/Cas9 induced PGE technique allows us to edit any individual letters in a given chapter without making any spelling errors [49]. In laboratories, the "CRISPR Craze" is quickly bridging the genotype and phenotype worlds in reverse genetics studies. It allows geneticists to scrutinize every building block of our genomes and introduce intentional alterations to our genetic materials that are otherwise only accessible to Mother Nature. In clinics, CRISPR/Cas9 holds promise to revolutionize the treatment of a wide range of incurable diseases and viruses such as cancer and HIV, and inherited disorders such as sickle-cell anemia, Down syndrome and Huntington's disease. This technique is hailed as a landmark in biological research among the most famous technological platforms such as PCR, RNAi and iPS cells, and a milestone in biomedical science [49, 97, 99].

Interestingly, CRISRP/Cas9 does not perform PGE by itself, but it only induces a targeted genomic lesion and invites the HDR pathways to introduce the desired modifications to the genome [12]. Although HDR is relatively well studied in yeast and human cells [6, 7, 14, 20, 37], the identity, property and hierarchy of the HDR pathways leading to the formation of PGE products remained elusive before this thesis research was conducted. Notably, PGE is not synonymous to HDR, but it specifically describes

the faction of HDR leading to the conversion of the desired knock-in mutations into the genome. Because not all the HDR pathways have the same propensity of inducing gene conversions, retrospectively, the composition and hierarchy of the PGE pathways may be very different from the natural HDR.

In this thesis research, I demonstrated that the meganuclease-induced PGE pathways in human cells are determined primarily by the types of homology donors and secondarily by the type and strandedness of genomic lesions **(Figure 1)**. dsDNA donors with a sizable heterology (>100bp) predominantly utilize the DSBR pathway both in the absence and presence of chromosomal DSBs, because it is the only long-tract gene conversion pathway that can efficiently convert large knock-in mutations without significantly sacrificing genomic stability. This pathway generates bidirectional conversion tracts with a linear distribution **(Figure 1, bottom left)**, which is centered at the central heterology being selected for during spontaneous PGE and at the chromosomal lesion during meganuclease-induced PGE. The average conversion tracts are exactly half of the lengths of the effective homology arms. Because we can simply increase the conversion tracts by using longer homology arms in the DSBR pathway, dsDNA donors are ideal for introducing large (multi-kilobase) mutations into the genome. The concept of dsDNA donors is not limited to linear and circular plasmids, but also applies to viruses with dsDNA PGE intermediates, such as adenoviruses and lentiviruses. Interestingly, I also prove that rAAV, as a single-strand DNA virus, becomes double-stranded during replication and serves as a type of dsDNA donor during spontaneous and meganuclease-induced PGE.

In contrast, single-strand ODN donors utilize the SDSA and ssDI pathways, respectively, depending on the strandedness of the genomic lesion that initiates PGE **(Figure 1, bottom right)**. When the effective genomic lesion is complementary to the

strand of the ODN donors, it initiates PGE via the SDSA pathway. This pathway generates unidirectional conversion tracts with a Gaussian distribution that starts about a few bp upstream of the genomic lesions; on the other hand, when the effective genomic lesion is same to the strand of the ODN donors, it initiates PGE via the ssDI pathway. This pathway produces bidirectional conversion tracts with a Gaussian distribution that is slightly biased towards the upstream region of the genomic lesion. Notably, both SDSA and ssDI are short-tract gene conversion pathways. The effective conversion tracts are about 30 bp long with different directional biases with respect to the effective genomic lesion. The Gaussian distribution of conversion tracts makes them extremely inefficient in converting large knock-in mutations used in dsDNA donors. In the presence of double-strand genomic lesions such as DSBs and paired nicks, the lesion complementary to the strand of the ODNs is preferentially utilized to initiate the SDSA pathway, as long as the knock-in mutation on the ODNs is within the effective conversion zone of this pathway. The ssDI pathway serves as a backup when SDSA fails to convert the knock-in mutation effectively.

Collectively, this work determined the identity, property and hierarchy of the PGE pathways in human cells using definitive molecular evidence. It did not only extend the paradigms of HDR-mediated gene conversion, but also established guidelines for the improvement of PGE:

1. ODN donors can efficiently convert knock-in mutations up to 30 bp, whereas larger modifications have to be converted using dsDNA donors.
2. dsDNA donors generate bidirectional conversion tracts with a linear distribution. Because of the mild drop-off on both direction of the curve, the knock-in mutations can be engineered into a wide range of regions (effective zones) as

long as there are enough flanking homology arms. The highest efficiency is achieved when the mutation is right at the position of the genomic lesion.

3. dsDNA donors mediate efficient PGE in the presence of DSBs and paired nicks, but not a single nick. The half life of dsDNA donors in human cells is better correlated with Cas9/sgRNA expression in DNA format.
4. ODN donors produce unidirectional and bidirectional conversion tracts with Gaussian distributions. The steep drop-offs in the conversion tracts restrict the effective conversion regions to a <30 bp zone. The predominant pathway and its conversion tracts have to be carefully considered to engineer the knock-in mutation into the effective zone.
5. ODN-mediated PGE can be efficiently induced by all kinds of genomic lesions. The half life of ODN donors is better correlated with the RNA or RNP format of Cas9/sgRNA expression *in vivo*.
6. Additional silent mutations have to be engineered to destroy the PAM sequence in the homology donors and prevent Cas9 from re-cutting the PGE products when PGE is induced by DSBs or paired nicks.
7. The anti-recombination activity of the mismatch repair system represents the biggest barrier of PGE in human cells. However, the level of suppression may vary in different PGE pathways depending on the length of heteroduplex DNA formed in the HDR intermediates. This suppression can be alleviated by knocking down of the mismatch repair genes or using homology donors with locked nucleic acids [184, 185].

Prospects

At the end of this work, I would like to propose three bold predictions about the trends in genetics and genomics studies. These predictions are not direct extrapolations from my current research, but the insights of the future genome engineering that I have drawn from more than 1500 publications in the field:

The timeline of genome engineering

Due to the unprecedented ease of genome engineering provided by the CRISPR/Cas9 technology, the timeline in generating genetically modified research models is infinitely approaching the life cycles of the specific organisms being used. Although the efficiency of genome engineering varies dramatically according to the DNA repair status of the specific organisms and the desired types of genomic modifications, multiplexed homozygous modifications can be routinely introduced in the F0 generation of the major model organisms [100-103], or within one round of genomic engineering in cultured human cells [167, 168]. Although the power of genome engineering has not penetrated every obscure model organism, transgenic mice with the desired genotypes can be generated within merely 4 weeks using embryo transfer [103], and in our hand, bi-allelic PGE can be achieved within 3 weeks in human cells. The future development of this technique will dramatically increase the list of genome-tractable organisms, and obviate the need of sequential genome engineering, genetic crossing, drug selection and Cre recombination for complex genome modifications in these organisms.

The shift of model organisms

The extreme ease in genome engineering may lead to a shift of the subjects of biological research from lower to higher model organisms. Historically, one of the

biggest reasons that lower model organisms were chosen for biological research was the existence of powerful tools to introduce desired genomic alterations via either forward or reverse genetics. Although the fundamental biological processes may be conserved from lower these organisms to human beings, there is no denying that lower organisms are different from human beings in many aspects. As a consequence, a significant amount of work has to be performed to verify the discoveries in the lower organisms before they can be extrapolated to human beings and make a realistic impact on human health. With the advancement of genome engineering technique and the ever-increasing list of genome tractable organisms, nevertheless, introducing the desired genomic modifications will no longer be a bottleneck in the future biological research. Thus model organisms will be chosen mainly based on pertinence to important biological questions and relevance to human health. It may lead to a dramatic shift of the subjects of biological research from lower organisms to higher models. In the future, traditional model organisms such as yeasts, flies, worms and fish may play a decreasing role in biological research, whereas transgenic mammalian models such as mice, pigs and monkeys may be more frequently employed in studies closely related to human health.

The flattening technology frontiers

The unprecedented ease in genome engineering and the global sharing of plasmids between institutions or through non-profit organizations (such as Addgene) may lead to the gradual disappearance of technological barriers between laboratories. As the tools of genome engineering are becoming more and more convenient and user-friendly, any scientists with basic laboratory trainings will be able to create the research models of the desired genotypes in the designated timeline of the published protocols. The edges between experienced laboratories and starters will become smaller and

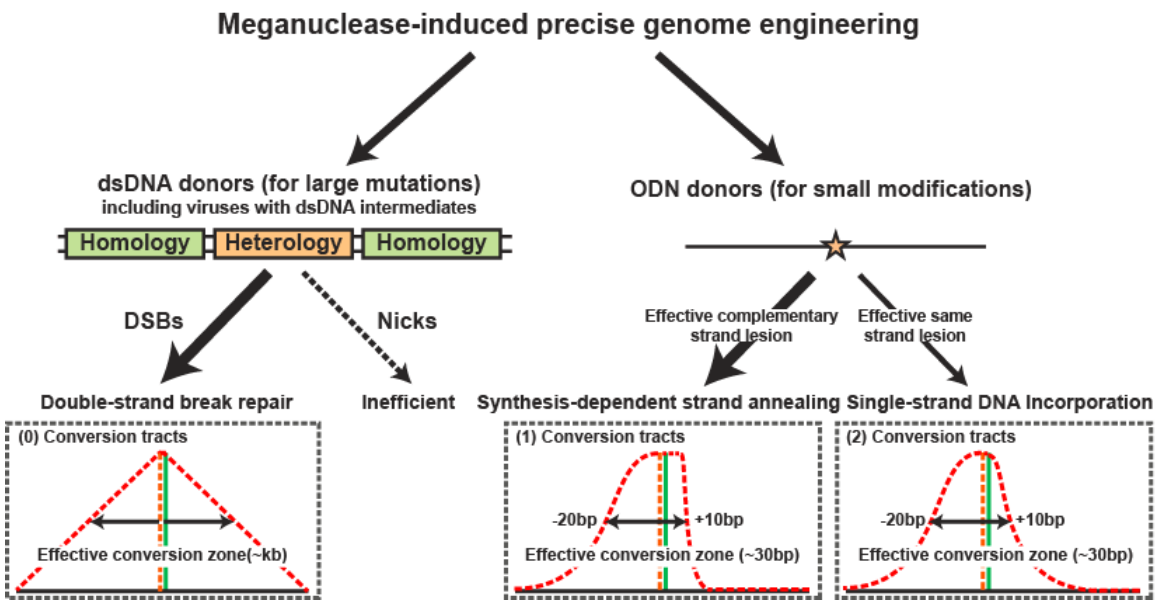
smaller, and eventually technology will no longer become the bottleneck of future biological research. One day, all laboratories will be starting at the same scratch line of the technology frontiers in the grand race of solving important biological questions. As the technological barriers become lower, the bar of publications will continue to rise. The winners will be the ones that can timely identify and solve important biological questions and continuously push forward the knowledge and technology frontiers. This will be the triumph of biology.

Figures and Legends

Figure 1. The hierarchy of meganuclease-induced HDR leading to PGE

PGE is defined as the fraction of HDR leading to the conversion of the desired knock-in mutations using exogenous homology donors. In the PGE products, retrospectively, the hierarchy of the HDR pathway is determined primarily by the types of homology donors and secondarily by the genomic lesions. dsDNA donors containing a sizable central heterology must be converted via the DSBR model, which generates long conversion tracts with a linear distribution. ODN donors can utilize both SDSA and ssDI pathways, depending on the strandedness of the ODNs and the relative position of the knock-in mutation to the genomic lesion. These pathways generate short conversion tracts in normal distributions. SDSA is preferentially utilized in the presence of compound genomic lesions when both pathways can convert the knock-in mutation effectively.

Figure 1


Meganuclease-induced precise genome engineering
dsDNA donors (for large mutations)
including viruses with dsDNA intermediates
Homology
Heterology
Homology
ODN donors (for small modifications)
DSBs
Nicks
Effective complementary strand lesion
Effective same strand lesion
Double-strand break repair
Inefficient
Synthesis-dependent strand annealing
Single-strand DNA Incorporation
(0) Conversion tracts
Effective conversion zone(~kb)
(1) Conversion tracts
-20bp
+10bp
Effective conversion zone (~30bp)
(2) Conversion tracts
-20bp
+10bp
Effective conversion zone (~30bp)

BIBLIOGRAPHY

1. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737-8. PubMed PMID: 13054692.
2. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. Ann Hum Biol. 2013;40(6):463-71. doi: 10.3109/03014460.2013.807878. PubMed PMID: 23829164.
3. van Gent DC, Hoeijmakers JH, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. Nat Rev Genet. 2001;2(3):196-206. doi: 10.1038/35056049. PubMed PMID: 11256071.
4. Moynahan ME. The cancer connection: BRCA1 and BRCA2 tumor suppression in mice and humans. Oncogene. 2002;21(58):8994-9007. doi: 10.1038/sj.onc.1206177. PubMed PMID: 12483515.
5. Jasin M. Homologous repair of DNA damage and tumorigenesis: the BRCA connection. Oncogene. 2002;21(58):8981-93. doi: 10.1038/sj.onc.1206176. PubMed PMID: 12483514.
6. Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol. 2006;7(10):739-50. doi: 10. 1038/nrm2008. PubMed PMID: 16926856.
7. San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. Annu Rev Biochem. 2008;77:229-57. Epub 2008/02/16. doi: 10.1146/annurev.biochem.77.061306.125255. PubMed PMID: 18275380.
8. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA repair. 2008;7(10):1765-71. doi: 10.1016/j.dnarep.2008.06.018. PubMed PMID: 18675941; PubMed Central PMCID: PMC2695993.
9. Betermier M, Bertrand P, Lopez BS. Is non-homologous end-joining really an inherently error-prone process? PLoS Genet. 2014;10(1):e1004086. doi: 10.1371/journal.pgen.1004086. PubMed PMID: 24453986; PubMed Central PMCID: PMC3894167.
10. Waters CA, Strande NT, Wyatt DW, Pryor JM, Ramsden DA. Nonhomologous end joining: a good solution for bad ends. DNA repair. 2014;17:39-51. doi: 10.1016/j.dnarep.2014.02.008. PubMed PMID: 24630899; PubMed Central PMCID:

PMC4024359.

11. Ginn SL, Alexander IE, Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to 2012 - an update. The journal of gene medicine. 2013;15(2):65-77. Epub 2013/01/29. doi: 10.1002/jgm.2698. PubMed PMID: 23355455.
12. Carroll D. Genome engineering with targetable nucleases. Annu Rev Biochem. 2014;83:409-39. doi: 10.1146/annurev-biochem-060713-035418. PubMed PMID: 24606144.
13. Crick F. The double helix: a personal view. Nature. 1974;248(5451):766-9. PubMed PMID: 4599081.
14. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annu Rev Genet. 2011;45:247-71. Epub 2011/09/14. doi: 10.1146/annurev-genet-110410-132435. PubMed PMID: 21910633.
15. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem. 2010;79:181-211. Epub 2010/03/03. doi: 10.1146/annurev.biochem.052308.093131. PubMed PMID: 20192759; PubMed Central PMCID: PMC3079308.
16. Mao Z, Bozzella M, Seluanov A, Gorbunova V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 2008;7(18):2902-6. PubMed PMID: 18769152; PubMed Central PMCID: PMC2754209.
17. Lombard DB, Chua KF, Mostoslavsky R, Franco S, Gostissa M, Alt FW. DNA repair, genome stability, and aging. Cell. 2005;120(4):497-512. doi: 10.1016/j.cell.2005.01.028. PubMed PMID: 15734682.
18. Game JC, Mortimer RK. A genetic study of x-ray sensitive mutants in yeast. Mutat Res. 1974;24(3):281-92. PubMed PMID: 4606119.
19. Li X, Heyer WD. Homologous recombination in DNA repair and DNA damage tolerance. Cell Res. 2008;18(1):99-113. doi: 10.1038/cr.2008.1. PubMed PMID: 18166982; PubMed Central PMCID: PMC3087377.
20. Heyer WD, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annu Rev Genet. 2010;44:113-39. Epub 2010/08/10. doi: 10.1146/annurev-genet-051710-150955. PubMed PMID: 20690856.
21. Paques F, Haber JE. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev. 1999;63(2):349-404.

PubMed PMID: 10357855; PubMed Central PMCID: PMC98970.

22. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A. 1981;78(10):6354-8. Epub 1981/10/01. PubMed PMID: 6273866; PubMed Central PMCID: PMCPMC349037.
23. Orr-Weaver TL, Szostak JW. Yeast recombination: the association between double-strand gap repair and crossing-over. Proc Natl Acad Sci U S A. 1983;80(14):4417-21. PubMed PMID: 6308623; PubMed Central PMCID: PMC384049.
24. R H. A mechanism for gene conversion in fungi. Genet Res. 1964;(5):282–304.
25. Meselson MS, Radding CM. A general model for genetic recombination. Proc Natl Acad Sci U S A. 1975;72(1):358-61. Epub 1975/01/01. PubMed PMID: 1054510; PubMed Central PMCID: PMCPMC432304.
26. Wu L, Hickson ID. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature. 2003;426(6968):870-4.
27. Swuec P, Costa A. Molecular mechanism of double Holliday junction dissolution. Cell Biosci. 2014;4:36. Epub 2014/07/26. doi: 10.1186/2045-3701-4-36. PubMed PMID: 25061510; PubMed Central PMCID: PMCPMC4109787.
28. Bizard AH, Hickson ID. The dissolution of double Holliday junctions. Cold Spring Harb Perspect Biol. 2014;6(7):a016477. Epub 2014/07/06. doi: 10.1101/cshperspect.a016477. PubMed PMID: 24984776.
29. Ferguson DO, Holloman WK. Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. Proc Natl Acad Sci U S A. 1996;93(11):5419-24. Epub 1996/05/28. PubMed PMID: 8643590; PubMed Central PMCID: PMCPMC39261.
30. Nassif N, Penney J, Pal S, Engels WR, Gloor GB. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol. 1994;14(3):1613-25. Epub 1994/03/01. PubMed PMID: 8114699; PubMed Central PMCID: PMCPMC358520.
31. Strathern JN, Klar AJ, Hicks JB, Abraham JA, Ivy JM, Nasmyth KA, et al. Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus. Cell. 1982;31(1):183-92. Epub 1982/11/01. PubMed PMID: 6297747.
32. Hastings PJ. Recombination in the eukaryotic nucleus. Bioessays. 1988;9(2-3):61-4. Epub 1988/08/01. doi: 10.1002/bies.950090206. PubMed PMID: 3066359.

33. Larocque JR, Jasin M. Mechanisms of recombination between diverged sequences in wild-type and BLM-deficient mouse and human cells. Mol Cell Biol. 2010;30(8):1887-97. Epub 2010/02/16. doi: 10.1128/MCB.01553-09. PubMed PMID: 20154148; PubMed Central PMCID: PMC2849462.
34. Borts RH, Haber JE. Meiotic recombination in yeast: alteration by multiple heterozygosities. Science. 1987;237(4821):1459-65. PubMed PMID: 2820060.
35. Llorente B, Smith CE, Symington LS. Break-induced replication: what is it and what is it for? Cell Cycle. 2008;7(7):859-64. PubMed PMID: 18414031.
36. Kraus E, Leung WY, Haber JE. Break-induced replication: a review and an example in budding yeast. Proc Natl Acad Sci U S A. 2001;98(15):8255-62. doi: 10.1073/pnas.151008198. PubMed PMID: 11459961; PubMed Central PMCID: PMC37429.
37. Symington LS. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol Mol Biol Rev. 2002;66(4):630-70, table of contents. Epub 2002/11/29. PubMed PMID: 12456786; PubMed Central PMCID: PMCPMC134659.
38. Radecke S, Radecke F, Peter I, Schwarz K. Physical incorporation of a single-stranded oligodeoxynucleotide during targeted repair of a human chromosomal locus. The journal of gene medicine. 2006;8(2):217-28. doi: 10.1002/jgm.828. PubMed PMID: 16142817.
39. Storici F, Snipe JR, Chan GK, Gordenin DA, Resnick MA. Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. Mol Cell Biol. 2006;26(20):7645-57. doi: 10.1128/MCB.00672-06. PubMed PMID: 16908537; PubMed Central PMCID: PMC1636868.
40. Davis L, Maizels N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. Proc Natl Acad Sci U S A. 2014;111(10):E924-32. doi: 10.1073/pnas.1400236111. PubMed PMID: 24556991; PubMed Central PMCID: PMC3956201.
41. Hendrie PC, Russell DW. Gene targeting with viral vectors. Mol Ther. 2005;12(1):9-17. Epub 2005/06/04. doi: 10.1016/j.ymthe.2005.04.006. PubMed PMID: 15932801.
42. Engelhardt JF. AAV hits the genomic bull's-eye. Nat Biotechnol. 2006;24(8):949-50. Epub 2006/08/11. doi: 10.1038/nbt0806-949. PubMed PMID: 16900138.
43. Vasileva A, Jessberger R. Precise hit: adeno-associated virus in gene targeting. Nat

Rev Microbiol. 2005;3(11):837-47. Epub 2005/11/02. doi: 10.1038/nrmicro1266. PubMed PMID: 16261169.

44. Keskin H, Shen Y, Huang F, Patel M, Yang T, Ashley K, et al. Transcript-RNA-templated DNA recombination and repair. Nature. 2014;515(7527):436-9. Epub 2014/09/05. doi: 10.1038/nature13682. PubMed PMID: 25186730.
45. Radhakrishnan SK, Jette N, Lees-Miller SP. Non-homologous end joining: emerging themes and unanswered questions. DNA repair. 2014;17:2-8. Epub 2014/03/04. doi: 10.1016/j.dnarep.2014.01.009. PubMed PMID: 24582502; PubMed Central PMCID: PMCPMC4084493.
46. Ma Y, Lu H, Schwarz K, Lieber MR. Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. Cell Cycle. 2005;4(9):1193-200. Epub 2005/08/06. PubMed PMID: 16082219.
47. Bunting SF, Nussenzweig A. End-joining, translocations and cancer. Nat Rev Cancer. 2013;13(7):443-54. Epub 2013/06/14. doi: 10.1038/nrc3537. PubMed PMID: 23760025.
48. Ghezraoui H, Piganeau M, Renouf B, Renaud JB, Sallmyr A, Ruis B, et al. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol Cell. 2014;55(6):829-42. Epub 2014/09/10. doi: 10.1016/j.molcel.2014.08.002. PubMed PMID: 25201414; PubMed Central PMCID: PMCPMC4398060.
49. Exclusive: 'Jaw-dropping' breakthrough hailed as landmark in fight against hereditary diseases as Crispr technique heralds genetic revolution The Independence. 2013.
50. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014;3:e04766. Epub 2014/12/17. doi: 10.7554/eLife.04766. PubMed PMID: 25497837; PubMed Central PMCID: PMCPMC4383097.
51. Miller DG, Petek LM, Russell DW. Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. Mol Cell Biol. 2003;23(10):3550-7. Epub 2003/05/02. PubMed PMID: 12724413; PubMed Central PMCID: PMC164770.
52. Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. Science. 2003;300(5620):763. Epub 2003/05/06. doi:

10.1126/science.1078395. PubMed PMID: 12730593.

53. Gellhaus K, Cornu TI, Heilbronn R, Cathomen T. Fate of recombinant adeno-associated viral vector genomes during DNA double-strand break-induced gene targeting in human cells. Hum Gene Ther. 2010;21(5):543-53. Epub 2009/12/22. doi: 10.1089/hum.2009.167. PubMed PMID: 20021219.
54. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 1996;93(3):1156-60. Epub 1996/02/06. PubMed PMID: 8577732; PubMed Central PMCID: PMCPMC40048.
55. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010;186(2):757-61. Epub 2010/07/28. doi: 10.1534/genetics.110.120717. PubMed PMID: 20660643; PubMed Central PMCID: PMCPMC2942870.
56. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816-21. Epub 2012/06/30. doi: 10.1126/science.1225829. PubMed PMID: 22745249.
57. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23. doi: 10.1126/science.1231143. PubMed PMID: 23287718; PubMed Central PMCID: PMC3795411.
58. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823-6. doi: 10.1126/science.1232033. PubMed PMID: 23287722; PubMed Central PMCID: PMC3712628.
59. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013;2:e00471. Epub 2013/02/07. doi: 10.7554/eLife.00471. PubMed PMID: 23386978; PubMed Central PMCID: PMCPMC3557905.
60. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31(3):230-2. Epub 2013/01/31. doi: 10.1038/nbt.2507. PubMed PMID: 23360966.
61. Hinnen A, Hicks JB, Fink GR. Transformation of yeast. Proc Natl Acad Sci U S A. 1978;75(4):1929-33. Epub 1978/04/01. PubMed PMID: 347451; PubMed Central

PMCID: PMCPMC392455.

62. Ilgen C, Farabaugh PJ, Hinnen A, Walsh J, Fink GR. Transformation of Yeast. In: Setlow J, Hollaender A, editors. Genetic Engineering. Genetic Engineering: Springer US; 1979. p. 117-32.
63. Orr-Weaver TL, Szostak JW, Rothstein RJ. Genetic applications of yeast transformation with linear and gapped plasmids. Methods Enzymol. 1983;101:228-45. Epub 1983/01/01. PubMed PMID: 6310326.
64. Langston LD, Symington LS. Gene targeting in yeast is initiated by two independent strand invasions. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(43):15392-7. Epub 2004/10/19. doi: 10.1073/pnas.0403748101. PubMed PMID: 15489271; PubMed Central PMCID: PMC524428.
65. Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature. 1985;317(6034):230-4. Epub 1985/09/19. PubMed PMID: 2995814.
66. Thomas KR, Folger KR, Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell. 1986;44(3):419-28. Epub 1986/02/14. PubMed PMID: 3002636.
67. Miyagawa K, Tsuruga T, Kinomura A, Usui K, Katsura M, Tashiro S, et al. A role for RAD54B in homologous recombination in human cells. The EMBO journal. 2002;21(1-2):175-80.
68. Yoshihara T, Ishida M, Kinomura A, Katsura M, Tsuruga T, Tashiro S, et al. XRCC3 deficiency results in a defect in recombination and increased endoreduplication in human cells. The EMBO journal. 2004;23(3):670-80.
69. Hiyama T, Katsura M, Yoshihara T, Ishida M, Kinomura A, Tonda T, et al. Haploinsufficiency of the Mus81-Eme1 endonuclease activates the intra-S-phase and G2/M checkpoints and promotes rereplication in human cells. Nucleic acids research. 2006;34(3):880-92.
70. Russell DW, Hirata RK. Human gene targeting by viral vectors. Nat Genet. 1998;18(4):325-30. Epub 1998/04/16. doi: 10.1038/ng0498-325. PubMed PMID: 9537413; PubMed Central PMCID: PMC3010411.
71. Kohli M, Rago C, Lengauer C, Kinzler KW, Vogelstein B. Facile methods for

generating human somatic cell gene knockouts using recombinant adeno-associated viruses. Nucleic acids research. 2004;32(1):e3.

72. Faruqi AF, Datta HJ, Carroll D, Seidman MM, Glazer PM. Triple-helix formation induces recombination in mammalian cells via a nucleotide excision repair-dependent pathway. Mol Cell Biol. 2000;20(3):990-1000. Epub 2000/01/11. PubMed PMID: 10629056; PubMed Central PMCID: PMCPMC85216.
73. Kuan JY, Glazer PM. Targeted gene modification using triplex-forming oligonucleotides. Methods Mol Biol. 2004;262:173-94. Epub 2004/02/11. doi: 10.1385/1-59259-761-0:173. PubMed PMID: 14769962.
74. Ferrara L, Kmiec EB. Camptothecin enhances the frequency of oligonucleotide-directed gene repair in mammalian cells by inducing DNA damage and activating homologous recombination. Nucleic acids research. 2004;32(17):5239-48. Epub 2004/10/07. doi: 10.1093/nar/gkh822. PubMed PMID: 15466591; PubMed Central PMCID: PMCPMC521643.
75. Wu XS, Xin L, Yin WX, Shang XY, Lu L, Watt RM, et al. Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. Proc Natl Acad Sci U S A. 2005;102(7):2508-13. Epub 2005/02/08. doi: 10.1073/pnas.0406991102. PubMed PMID: 15695590; PubMed Central PMCID: PMCPMC548982.
76. Huen MS, Li XT, Lu LY, Watt RM, Liu DP, Huang JD. The involvement of replication in single stranded oligonucleotide-mediated gene repair. Nucleic acids research. 2006;34(21):6183-94. Epub 2006/11/08. doi: 10.1093/nar/gkl852. PubMed PMID: 17088285; PubMed Central PMCID: PMCPMC1693898.
77. Suzuki T. Targeted gene modification by oligonucleotides and small DNA fragments in eukaryotes. Front Biosci. 2008;13:737-44. Epub 2007/11/06. PubMed PMID: 17981583.
78. Jensen NM, Dalsgaard T, Jakobsen M, Nielsen RR, Sorensen CB, Bolund L, et al. An update on targeted gene repair in mammalian cells: methods and mechanisms. J Biomed Sci. 2011;18:10. Epub 2011/02/03. doi: 10.1186/1423-0127-18-10. PubMed PMID: 21284895; PubMed Central PMCID: PMCPMC3042377.
79. Radecke F, Peter I, Radecke S, Gellhaus K, Schwarz K, Cathomen T. Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break. Mol Ther.

2006;14(6):798-808. doi: 10.1016/j.ymthe.2006.06.008. PubMed PMID: 16904944.

80. Davis L, Maizels N. DNA nicks promote efficient and safe targeted gene correction. PloS one. 2011;6(9):e23981. doi: 10.1371/journal.pone.0023981. PubMed PMID: 21912657; PubMed Central PMCID: PMC3164693.
81. Barzel A, Privman E, Peeri M, Naor A, Shachar E, Burstein D, et al. Native homing endonucleases can target conserved genes in humans and in animal models. Nucleic acids research. 2011;39(15):6646-59. Epub 2011/04/29. doi: 10.1093/nar/gkr242. PubMed PMID: 21525128; PubMed Central PMCID: PMCPMC3159444.
82. Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, Gloor GB, et al. Tapping natural reservoirs of homing endonucleases for targeted gene modification. Proc Natl Acad Sci U S A. 2011;108(32):13077-82. Epub 2011/07/26. doi: 10.1073/pnas.1107719108. PubMed PMID: 21784983; PubMed Central PMCID: PMCPMC3156218.
83. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 1985;4(6):1609-14. Epub 1985/06/01. PubMed PMID: 4040853; PubMed Central PMCID: PMCPMC554390.
84. Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic acids research. 2000;28(17):3361-9. Epub 2000/08/23. PubMed PMID: 10954606; PubMed Central PMCID: PMCPMC110700.
85. Carroll D. Genome engineering with zinc-finger nucleases. Genetics. 2011;188(4):773-82. Epub 2011/08/11. doi: 10.1534/genetics.111.131433. PubMed PMID: 21828278; PubMed Central PMCID: PMC3176093.
86. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. Science. 2011;333(6051):1843-6. Epub 2011/10/01. doi: 10.1126/science.1204094. PubMed PMID: 21960622.
87. Cermak T, Starker CG, Voytas DF. Efficient design and assembly of custom TALENs using the Golden Gate platform. Methods Mol Biol. 2015;1239:133-59. Epub 2014/11/20. doi: 10.1007/978-1-4939-1862-1_7. PubMed PMID: 25408404.
88. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem. 2013;82:237-66. Epub 2013/03/19. doi: 10.1146/annurev-biochem-072911-172315. PubMed PMID:

23495939.

89. Hynes AP, Villion M, Moineau S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat Commun. 2014;5:4399. Epub 2014/07/25. doi: 10.1038/ncomms5399. PubMed PMID: 25056268.
90. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol. 2008;190(4):1390-400. Epub 2007/12/11. doi: 10.1128/jb.01412-07. PubMed PMID: 18065545; PubMed Central PMCID: PMCPMC2238228.
91. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A. 2011;108(25):10098-103. Epub 2011/06/08. doi: 10.1073/pnas.1104144108. PubMed PMID: 21646539; PubMed Central PMCID: PMCPMC3121866.
92. Rutkauskas M, Sinkunas T, Songailiene I, Tikhomirova MS, Siksnys V, Seidel R. Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell Rep. 2015. Epub 2015/03/11. doi: 10.1016/j.celrep.2015.01.067. PubMed PMID: 25753419.
93. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci U S A. 2014;111(27):9798-803. Epub 2014/06/10. doi: 10.1073/pnas.1402597111. PubMed PMID: 24912165; PubMed Central PMCID: PMCPMC4103346.
94. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. Nat Methods. 2013;10(10):957-63. Epub 2013/10/01. doi: 10.1038/nmeth.2649. PubMed PMID: 24076990; PubMed Central PMCID: PMCPMC4051438.
95. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014;157(6):1262-78. doi: 10.1016/j.cell.2014.05.010. PubMed PMID: 24906146; PubMed Central PMCID: PMC4343198.
96. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346(6213):1258096. doi: 10.1126/science.1258096. PubMed PMID: 25430774.
97. Pennisi E. The CRISPR craze. Science. 2013;341(6148):833-6. doi: 10.1126/science.341.6148.833. PubMed PMID: 23970676.

98. Baker M. Gene editing at CRISPR speed. Nat Biotechnol. 2014;32(4):309-12. Epub 2014/04/10. doi: 10.1038/nbt.2863. PubMed PMID: 24714470.
99. Ledford H. CRISPR, the disruptor. Nature. 2015;522(7554):20-4. Epub 2015/06/05. doi: 10.1038/522020a. PubMed PMID: 26040877.
100. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research. 2013;41(7):4336-43. Epub 2013/03/06. doi: 10.1093/nar/gkt135. PubMed PMID: 23460208; PubMed Central PMCID: PMCPMC3627607.
101. Bassett AR, Tibbit C, Ponting CP, Liu JL. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep. 2013;4(1):220-8. Epub 2013/07/06. doi: 10.1016/j.celrep.2013.06.020. PubMed PMID: 23827738; PubMed Central PMCID: PMCPMC3714591.
102. Liu P, Long L, Xiong K, Yu B, Chang N, Xiong JW, et al. Heritable/conditional genome editing in C. elegans using a CRISPR-Cas9 feeding system. Cell Res. 2014;24(7):886-9. Epub 2014/05/31. doi: 10.1038/cr.2014.73. PubMed PMID: 24874953; PubMed Central PMCID: PMCPMC4085767.
103. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat Protoc. 2014;9(8):1956-68. Epub 2014/07/25. doi: 10.1038/nprot.2014.134. PubMed PMID: 25058643.
104. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977-9. Epub 2013/07/31. doi: 10.1038/nmeth.2598. PubMed PMID: 23892898; PubMed Central PMCID: PMCPMC3794058.
105. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013;10(10):973-6. Epub 2013/07/31. doi: 10.1038/nmeth.2600. PubMed PMID: 23892895; PubMed Central PMCID: PMCPMC3911785.
106. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, et al. Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013;500(7463):472-6. Epub 2013/07/24. doi: 10.1038/nature12466. PubMed PMID: 23877069; PubMed Central PMCID: PMCPMC3856241.
107. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell.

2014;159(3):635-46. Epub 2014/10/14. doi: 10.1016/j.cell.2014.09.039. PubMed PMID: 25307933; PubMed Central PMCID: PMCPMC4252608.

108. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. Science. 2014;343(6166):80-4. Epub 2013/12/18. doi: 10.1126/science.1246981. PubMed PMID: 24336569; PubMed Central PMCID: PMCPMC3972032.

109. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343(6166):84-7. Epub 2013/12/18. doi: 10.1126/science.1247005. PubMed PMID: 24336571; PubMed Central PMCID: PMCPMC4089965.

110. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell. 2015;160(6):1246-60. Epub 2015/03/10. doi: 10.1016/j.cell.2015.02.038. PubMed PMID: 25748654; PubMed Central PMCID: PMCPMC4380877.

111. Li J, Stoddard TJ, Demorest ZL, Lavoie PO, Luo S, Clasen BM, et al. Multiplexed, targeted gene editing in Nicotiana benthamiana for glyco-engineering and monoclonal antibody production. Plant Biotechnol J. 2015. Epub 2015/05/27. doi: 10.1111/pbi.12403. PubMed PMID: 26011187.

112. Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, et al. Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife. 2014;3. Epub 2014/08/21. doi: 10.7554/eLife.03703. PubMed PMID: 25139909; PubMed Central PMCID: PMCPMC4161972.

113. Brophy JA, Voigt CA. Principles of genetic circuit design. Nat Methods. 2014;11(5):508-20. Epub 2014/05/02. doi: 10.1038/nmeth.2926. PubMed PMID: 24781324; PubMed Central PMCID: PMCPMC4230274.

114. Hu W, Kaminski R, Yang F, Zhang Y, Cosentino L, Li F, et al. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. Proc Natl Acad Sci U S A. 2014;111(31):11461-6. Epub 2014/07/23. doi: 10.1073/pnas.1405186111. PubMed PMID: 25049410; PubMed Central PMCID: PMCPMC4128125.

115. Galizi R, Doyle LA, Menichelli M, Bernardini F, Deredec A, Burt A, et al. A synthetic sex ratio distortion system for the control of the human malaria mosquito. Nat Commun. 2014;5:3977. Epub 2014/06/11. doi: 10.1038/ncomms4977. PubMed

PMID: 24915045; PubMed Central PMCID: PMCPMC4057611.

116. Gantz VM, Bier E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. Science. 2015;348(6233):442-4. Epub 2015/04/25. doi: 10.1126/science.aaa5945. PubMed PMID: 25908821.
117. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. 2014;15(12):554. Epub 2014/12/06. doi: 10.1186/s13059-014-0554-4. PubMed PMID: 25476604; PubMed Central PMCID: PMCPMC4290824.
118. Long C, McAnally JR, Shelton JM, Mireault AA, Bassel-Duby R, Olson EN. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science. 2014;345(6201):1184-8. Epub 2014/08/16. doi: 10.1126/science.1254445. PubMed PMID: 25123483; PubMed Central PMCID: PMCPMC4398027.
119. Liang P, Xu Y, Zhang X, Ding C, Huang R, Zhang Z, et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell. 2015;6(5):363-72. Epub 2015/04/22. doi: 10.1007/s13238-015-0153-5. PubMed PMID: 25894090; PubMed Central PMCID: PMCPMC4417674.
120. Hendrickson EA. Gene targeting in human somatic cells. In: Conn PM, editor. Source Book of Models for Biomedical Research. Totowa, NJ: Humana Press, Inc.; 2008. p. 509-25.
121. Forget AL, Kowalczykowski SC. Single-molecule imaging brings Rad51 nucleoprotein filaments into focus. Trends Cell Biol. 2010;20(5):269-76. Epub 2010/03/20. doi: 10.1016/j.tcb.2010.02.004. PubMed PMID: 20299221; PubMed Central PMCID: PMC2862779.
122. Hastings PJ, McGill C, Shafer B, Strathern JN. Ends-in vs. ends-out recombination in yeast. Genetics. 1993;135(4):973-80. Epub 1993/12/01. PubMed PMID: 8307337; PubMed Central PMCID: PMC1205758.
123. Mitchel K, Zhang H, Welz-Voegele C, Jinks-Robertson S. Molecular structures of crossover and noncrossover intermediates during gap repair in yeast: implications for recombination. Mol Cell. 2010;38(2):211-22. Epub 2010/04/27. doi: 10.1016/j.molcel.2010.02.028. PubMed PMID: 20417600; PubMed Central PMCID: PMC2865147.

124. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell. 1987;51(3):503-12. Epub 1987/11/06. PubMed PMID: 2822260.
125. Li J, Baker MD. Mechanisms involved in targeted gene replacement in mammalian cells. Genetics. 2000;156(2):809-21. Epub 2000/10/03. PubMed PMID: 11014826; PubMed Central PMCID: PMC1461283.
126. Li J, Read LR, Baker MD. The mechanism of mammalian gene replacement is consistent with the formation of long regions of heteroduplex DNA associated with two crossing-over events. Mol Cell Biol. 2001;21(2):501-10. Epub 2001/01/03. doi: 10.1128/MCB.21.2.501-510.2001. PubMed PMID: 11134338; PubMed Central PMCID: PMC86609.
127. McCulloch RD, Baker MD. Analysis of one-sided marker segregation patterns resulting from mammalian gene targeting. Genetics. 2006;172(3):1767-81. Epub 2006/03/24. doi: 10.1534/genetics.105.051680. PubMed PMID: 16554412; PubMed Central PMCID: PMC1456313.
128. Ruksc A, Bell-Rogers PL, Smith JD, Baker MD. Analysis of spontaneous gene conversion tracts within and between mammalian chromosomes. Journal of molecular biology. 2008;377(2):337-51. Epub 2008/02/12. doi: 10.1016/j.jmb.2008.01.036. PubMed PMID: 18262541.
129. Khan IF, Hirata RK, Russell DW. AAV-mediated gene targeting methods for human cells. Nat Protoc. 2011;6(4):482-501. Epub 2011/04/02. doi: 10.1038/nprot.2011.301. PubMed PMID: 21455185.
130. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. Cell. 1983;33(1):25-35. Epub 1983/05/01. PubMed PMID: 6380756.
131. Hilliker AJ, Harauz G, Reaume AG, Gray M, Clark SH, Chovnick A. Meiotic gene conversion tract length distribution within the rosy locus of Drosophila melanogaster. Genetics. 1994;137(4):1019-26. Epub 1994/08/01. PubMed PMID: 7982556; PubMed Central PMCID: PMC1206049.
132. Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M. Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol. 1998;18(1):93-101. Epub 1998/01/07. PubMed PMID: 9418857; PubMed Central PMCID: PMC121458.

133. de Massy B. Distribution of meiotic recombination sites. Trends Genet. 2003;19(9):514-22. Epub 2003/09/06. doi: 10.1016/S0168-9525(03)00201-4. PubMed PMID: 12957545.
134. Stark JM, Pierce AJ, Oh J, Pastink A, Jasin M. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. Mol Cell Biol. 2004;24(21):9305-16. Epub 2004/10/16. doi: 10.1128/MCB.24.21.9305-9316.2004. PubMed PMID: 15485900; PubMed Central PMCID: PMC522275.
135. Thomas KR, Capecchi MR. Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. Nature. 1986;324(6092):34-8. Epub 1986/11/06. doi: 10.1038/324034a0. PubMed PMID: 3785372.
136. Kirkpatrick DT, Petes TD. Repair of DNA loops involves DNA-mismatch and nucleotide-excision repair proteins. Nature. 1997;387(6636):929-31. Epub 1997/06/26. doi: 10.1038/43225. PubMed PMID: 9202128.
137. Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, et al. Mutation of a mutL homolog in hereditary colon cancer. Science. 1994;263(5153):1625-9. Epub 1994/03/18. PubMed PMID: 8128251.
138. Yabuta T, Shinmura K, Yamane A, Yamaguchi S, Takenoshita S, Yokota J. Effect of exogenous MSH6 and POLD1 expression on the mutation rate of the HPRT locus in a human colon cancer cell line with mutator phenotype, DLD-1. Int J Oncol. 2004;24(3):697-702. Epub 2004/02/10. PubMed PMID: 14767555.
139. Miller DG, Trobridge GD, Petek LM, Jacobs MA, Kaul R, Russell DW. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J Virol. 2005;79(17):11434-42. Epub 2005/08/17. doi: 10.1128/JVI.79.17.11434-11442.2005. PubMed PMID: 16103194; PubMed Central PMCID: PMC1193581.
140. Nakai H, Wu X, Fuess S, Storm TA, Munroe D, Montini E, et al. Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. J Virol. 2005;79(6):3606-14. Epub 2005/02/26. doi: 10.1128/JVI.79.6.3606-3614.2005. PubMed PMID: 15731255; PubMed Central PMCID: PMC1075691.
141. Janovitz T, Klein IA, Oliveira T, Mukherjee P, Nussenzweig MC, Sadelain M, et al. High-throughput sequencing reveals principles of Adeno-Associated Virus Serotype 2 integration. J Virol. 2013;in press. Epub 2013/05/31. doi: 10.1128/JVI.01135-13. PubMed PMID: 23720718.

142. Goncalves MA. Adeno-associated virus: from defective virus to effective vector. Virol J. 2005;2:43. Epub 2005/05/10. doi: 10.1186/1743-422X-2-43. PubMed PMID: 15877812; PubMed Central PMCID: PMC1131931.
143. Villalobos MJ, Betti CJ, Vaughan AT. Detection of DNA double-strand breaks and chromosome translocations using ligation-mediated PCR and inverse PCR. Methods Mol Biol. 2006;314:109-21. Epub 2006/05/06. doi: 10.1385/1-59259-973-7:109. PubMed PMID: 16673878.
144. Hollingsworth NM, Brill SJ. The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. Genes Dev. 2004;18(2):117-25. Epub 2004/01/31. doi: 10.1101/gad.1165904. PubMed PMID: 14752007; PubMed Central PMCID: PMC1851908.
145. Svendsen JM, Harper JW. GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. Genes Dev. 2010;24(6):521-36. Epub 2010/03/06. doi: 10.1101/gad.1903510. PubMed PMID: 20203129; PubMed Central PMCID: PMC2841330.
146. Wright DA, Townsend JA, Winfrey RJ, Jr., Irwin PA, Rajagopal J, Lonosky PM, et al. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J. 2005;44(4):693-705. Epub 2005/11/03. doi: 10.1111/j.1365-313X.2005.02551.x. PubMed PMID: 16262717.
147. Andersen SL, Sekelsky J. Meiotic versus mitotic recombination: two different routes for double-strand break repair: the different functions of meiotic versus mitotic DSB repair are reflected in different pathway usage and different outcomes. Bioessays. 2010;32(12):1058-66. Epub 2010/10/23. doi: 10.1002/bies.201000087. PubMed PMID: 20967781; PubMed Central PMCID: PMC3090628.
148. Umar A, Boyer JC, Kunkel TA. DNA loop repair by human cell extracts. Science. 1994;266(5186):814-6. Epub 1994/11/04. PubMed PMID: 7973637.
149. Harfe BD, Jinks-Robertson S. Mismatch repair proteins and mitotic genome stability. Mutat Res. 2000;451(1-2):151-67. Epub 2000/08/01. PubMed PMID: 10915870.
150. Siehler SY, Schrauder M, Gerischer U, Cantor S, Marra G, Wiesmuller L. Human MutL-complexes monitor homologous recombination independently of mismatch repair. DNA repair. 2009;8(2):242-52. Epub 2008/11/22. doi: 10.1016/j.dnarep.2008.10.011. PubMed PMID: 19022408; PubMed Central PMCID: PMC2909591.

151. Hendrie PC, Hirata RK, Russell DW. Chromosomal integration and homologous gene targeting by replication-incompetent vectors based on the autonomous parvovirus minute virus of mice. J Virol. 2003;77(24):13136-45. Epub 2003/12/04. PubMed PMID: 14645570; PubMed Central PMCID: PMC296056.
152. Summerford C, Samulski RJ. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. J Virol. 1998;72(2):1438-45. Epub 1998/01/28. PubMed PMID: 9445046; PubMed Central PMCID: PMC124624.
153. McCarty DM, Young SM, Jr., Samulski RJ. Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu Rev Genet. 2004;38:819-45. Epub 2004/12/01. doi: 10.1146/annurev.genet.37.110801.143717. PubMed PMID: 15568995.
154. Cataldi MP, McCarty DM. Hairpin-end conformation of adeno-associated virus genome determines interactions with DNA-repair pathways. Gene Ther. 2013;20(6):686-93. Epub 2012/11/16. doi: 10.1038/gt.2012.86. PubMed PMID: 23151519; PubMed Central PMCID: PMC3578132.
155. Deng C, Capecchi MR. 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(8):3365-71. Epub 1992/08/01. PubMed PMID: 1321331; PubMed Central PMCID: PMC364584.
156. Shen P, Huang HV. Homologous recombination in Escherichia coli: dependence on substrate length and homology. Genetics. 1986;112(3):441-57. Epub 1986/03/01. PubMed PMID: 3007275; PubMed Central PMCID: PMC1202756.
157. Fishman-Lobell J, Haber JE. Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. Science. 1992;258(5081):480-4. Epub 1992/10/16. PubMed PMID: 1411547.
158. Al-Minawi AZ, Saleh-Gohari N, Helleday T. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. Nucleic acids research. 2008;36(1):1-9. Epub 2007/10/27. doi: 10.1093/nar/gkm888. PubMed PMID: 17962301; PubMed Central PMCID: PMC2248766.
159. Murayama Y, Kurokawa Y, Mayanagi K, Iwasaki H. Formation and branch migration of Holliday junctions mediated by eukaryotic recombinases. Nature. 2008;451(7181):1018-21. Epub 2008/02/08. doi: 10.1038/nature06609. PubMed

PMID: 18256600.

160. Choulika A, Perrin A, Dujon B, Nicolas JF. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol Cell Biol. 1995;15(4):1968-73. Epub 1995/04/01. PubMed PMID: 7891691; PubMed Central PMCID: PMC230423.
161. Aarts M, te Riele H. Progress and prospects: oligonucleotide-directed gene modification in mouse embryonic stem cells: a route to therapeutic application. Gene Ther. 2011;18(3):213-9. Epub 2010/12/17. doi: 10.1038/gt.2010.161. PubMed PMID: 21160530.
162. Charpentier E, Doudna JA. Biotechnology: Rewriting a genome. Nature. 2013;495(7439):50-1. Epub 2013/03/08. doi: 10.1038/495050a. PubMed PMID: 23467164.
163. Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods. 2011;8(9):753-5. Epub 2011/07/19. doi: 10.1038/nmeth.1653. PubMed PMID: 21765410; PubMed Central PMCID: PMC3617923.
164. Bedell VM, Wang Y, Campbell JM, Poshusta TL, Starker CG, Krug RG, 2nd, et al. In vivo genome editing using a high-efficiency TALEN system. Nature. 2012;491(7422):114-8. Epub 2012/09/25. doi: 10.1038/nature11537. PubMed PMID: 23000899; PubMed Central PMCID: PMC3491146.
165. Thaler DS, Stahl FW. DNA double-chain breaks in recombination of phage lambda and of yeast. Annu Rev Genet. 1988;22:169-97. Epub 1988/01/01. doi: 10.1146/annurev.ge.22.120188.001125. PubMed PMID: 2977087.
166. Rahn JJ, Rowley B, Lowery MP, Coletta LD, Limanni T, Nairn RS, et al. Effects of varying gene targeting parameters on processing of recombination intermediates by ERCC1-XPF. DNA repair. 2011;10(2):188-98. Epub 2010/12/03. doi: 10.1016/j.dnarep.2010.10.011. PubMed PMID: 21123118; PubMed Central PMCID: PMC3034777.
167. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281-308. Epub 2013/10/26. doi: 10.1038/nprot.2013.143. PubMed PMID: 24157548; PubMed Central PMCID: PMCPMC3969860.
168. Byrne SM, Mali P, Church GM. Genome editing in human stem cells. Methods

Enzymol. 2014;546:119-38. Epub 2014/11/16. doi: 10.1016/b978-0-12-801185-0.00006-4. PubMed PMID: 25398338; PubMed Central PMCID: PMCPMC4408990.

169. Kan Y, Ruis B, Lin S, Hendrickson EA. The mechanism of gene targeting in human somatic cells. PLoS Genet. 2014;10(4):e1004251. doi: 10.1371/journal.pgen.1004251. PubMed PMID: 24699519; PubMed Central PMCID: PMC3974634.

170. Parekh-Olmedo H, Ferrara L, Brachman E, Kmiec EB. Gene therapy progress and prospects: targeted gene repair. Gene Ther. 2005;12(8):639-46. doi: 10.1038/sj.gt.3302511. PubMed PMID: 15815682.

171. Katada H, Ren Y, Shigi N, Komiyama M. Site-specific gene manipulation of fluorescent proteins using artificial restriction DNA cutter. Nucleic acids symposium series. 2008;(52):483-4. doi: 10.1093/nass/nrn245. PubMed PMID: 18776464.

172. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154(6):1380-9. doi: 10.1016/j.cell.2013.08.021. PubMed PMID: 23992846; PubMed Central PMCID: PMC3856256.

173. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 2014;24(1):132-41. Epub 2013/11/21. doi: 10.1101/gr.162339.113. PubMed PMID: 24253446; PubMed Central PMCID: PMCPMC3875854.

174. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods. 2014;11(4):399-402. doi: 10.1038/nmeth.2857. PubMed PMID: 24584192.

175. Balakrishnan L, Bambara RA. Flap endonuclease 1. Annu Rev Biochem. 2013;82:119-38. Epub 2013/03/05. doi: 10.1146/annurev-biochem-072511-122603. PubMed PMID: 23451868; PubMed Central PMCID: PMCPMC3679248.

176. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annual review of genetics. 2011;45:247-71. doi: 10.1146/annurev-genet-110410-132435. PubMed PMID: 21910633.

177. Ramirez CL, Certo MT, Mussolino C, Goodwin MJ, Cradick TJ, McCaffrey AP, et al. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucleic acids research. 2012;40(12):5560-8. Epub 2012/03/01. doi: 10.1093/nar/gks179. PubMed PMID: 22373919; PubMed Central PMCID:

PMCPMC3384306.

178. Heyer WD, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annual review of genetics. 2010;44:113-39. doi: 10.1146/annurev-genet-051710-150955. PubMed PMID: 20690856; PubMed Central PMCID: PMC4114321.

179. Langston LD, Symington LS. Gene targeting in yeast is initiated by two independent strand invasions. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(43):15392-7. doi: 10.1073/pnas.0403748101. PubMed PMID: 15489271; PubMed Central PMCID: PMC524428.

180. Larocque JR, Jasin M. Mechanisms of recombination between diverged sequences in wild-type and BLM-deficient mouse and human cells. Molecular and cellular biology. 2010;30(8):1887-97. doi: 10.1128/MCB.01553-09. PubMed PMID: 20154148; PubMed Central PMCID: PMC2849462.

181. Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M. Gene conversion tracts from double-strand break repair in mammalian cells. Molecular and cellular biology. 1998;18(1):93-101. PubMed PMID: 9418857; PubMed Central PMCID: PMC121458.

182. Duda K, Lonowski LA, Kofoed-Nielsen M, Ibarra A, Delay CM, Kang Q, et al. High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. Nucleic acids research. 2014;42(10):e84. Epub 2014/04/23. doi: 10.1093/nar/gku251. PubMed PMID: 24753413; PubMed Central PMCID: PMCPMC4041425.

183. Khan IF, Hirata RK, Russell DW. AAV-mediated gene targeting methods for human cells. Nature protocols. 2011;6(4):482-501. doi: 10.1038/nprot.2011.301. PubMed PMID: 21455185; PubMed Central PMCID: PMC3739714.

184. Parekh-Olmedo H, Drury M, Kmiec EB. Targeted nucleotide exchange in Saccharomyces cerevisiae directed by short oligonucleotides containing locked nucleic acids. Chem Biol. 2002;9(10):1073-84. Epub 2002/10/29. PubMed PMID: 12401492.

185. Andrieu-Soler C, Casas M, Faussat AM, Gandolphe C, Doat M, Tempe D, et al. Stable transmission of targeted gene modification using single-stranded oligonucleotides with flanking LNAs. Nucleic acids research. 2005;33(12):3733-42. Epub 2005/07/09. doi: 10.1093/nar/gki686. PubMed PMID: 16002788; PubMed

Central PMCID: PMCPMC1174897.