Precision Genome Editing in the CRISPR Era
Precision Genome Editing in the CRISPR Era

Jayme Salsman and Graham Dellaire

Department of Pathology, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4R2

Departments of Pathology and Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4R2; Beatrice Hunter Cancer Research Institute, Halifax, Nova Scotia, B3H 4R2, Canada

* To whom correspondence should be addressed. Tel: (902) 494-4730; Fax: (902) 494-2519; Email: dellaire@dal.ca

KEYWORDS: CRISPR, Cas9, Genome Editing, DNA repair, Gene therapy

ABSTRACT:

With the introduction of precision genome editing using CRISPR/Cas9 technology, we have entered a new era of genetic engineering and gene therapy. With RNA-guided endonucleases, such as Cas9, it is possible to engineer DNA double strand breaks (DSB) at specific genomic loci. DSB repair by the error-prone non-homologous end joining (NHEJ) pathway can disrupt a target gene by generating insertions and deletions. Alternatively, Cas9-mediated DSBs can be repaired by homology directed repair (HDR) using a homologous DNA repair template, thus allowing precise gene editing by incorporating genetic changes into the repair template. HDR can introduce gene sequences for protein epitope tags, delete genes, make point mutations or alter enhancer and promoter activities. In anticipation of adapting this technology for gene therapy in human somatic cells, much focus has been placed on increasing the fidelity of CRISPR/Cas9 and increasing HDR efficiency to improve precision genome editing. In this review, we will discuss applications of CRISPR technology for gene inactivation and genome editing with a focus on approaches to enhancing CRISPR/Cas9-mediated HDR for the generation of cell and animal models, and conclude with a discussion of recent advances and challenges towards the application of this technology for gene therapy in humans.
1. History of Genome Editing

DNA damage in the form of DNA double strand breaks (DSBs) can occur after exposure to ionizing radiation and DNA damaging chemotherapy, during DNA replication, or experimentally through the action of endonucleases (endogenous or exogenously introduced) (Wyman and Kanaar 2006). DNA DSBs are often deleterious to the cell, resulting in genome instability and disease-causing mutations. The DNA repair machinery helps to not only repair damage, but is also involved in important biological processes such as meiotic recombination, antibody class switching and VDJ rearrangements (Stavnezer and Schrader 2014, Roth 2014, Hunter 2015). The ability to harness DNA damage and repair responses forms the basis of gene editing. There are three major repair pathways for DSBs known as canonical non-homologous end joining (c-NHEJ), alternative NHEJ (alt-NHEJ) and homology recombination (HR) (Fig 1) (Ceccaldi, Rondinelli and D'Andrea 2016). The NHEJ pathways are active throughout the cell cycle and are therefore the most common mechanism for DSB repair even though they are error-prone and often result in insertions and deletions (indels) (Chiruvella, Liang and Wilson 2013). HR, is initiated in S- or G2-phase of the cell cycle when a sister chromatid is available to provide a donor homology template for repair by homologous recombination (Mladenov et al. 2016, Branzei and Foiani 2008); a process referred to as homology-directed repair (HDR) in the context of an exogenous DNA repair template (Fig 1). This property of HR was used in the early 1980s to both insert and to repair genes in mammalian cells by HDR (Smithies et al. 1985, Thomas and Capecchi 1987). These techniques were then applied to mouse embryos to generate transgenic mice, and the modern era of gene engineering was born (Capecchi 2005). These early seminal studies were carried out by the laboratories of Drs. Oliver Smithies, Mario Capecchi and Martin Evens, who together pioneered HR-mediated gene editing in mouse embryos and would later share the 2007 Nobel Prize in Physiology or Medicine “for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells” (Vogel 2007).

In the 1990s, gene targeting in mammalian cells using HR was further enhanced by Dr. Maria Jasin and colleagues by using yeast homing endonuclease I-SceI (Rouet, Smih and Jasin 1994, Smih et al. 1995). The SceI enzyme generates site-specific breaks in chromatin carrying an ectopic copy of the 18-bp
rare restriction site for this enzyme, a site that does not occur in mouse or human genomes and thus is highly specific. Cutting with the SceI enzyme could enhance HR up to 500 fold at the target locus, and these studies would foreshadow the use of designer site-specific endonucleases to enhance gene editing. These included the zinc finger nucleases (ZFNs) and the Transcription Activator-Like Effector Nucleases (TALENs), which combined the DNA binding specificity of zinc fingers or TALE transcription factors from plants with the FokI endonuclease DNA cutting activity (Miller et al. 2011, Kim, Cha and Chandrasegaran 1996). However, the use of both ZFNs and TALENs had several caveats. These included issues with limited targeting capability due to the available ZFN libraries, TALENs could be expensive and time consuming to construct and test, and vectors encoding TALENs are often too large or complex for genome-wide screening or viral delivery as would be required for gene therapies. These challenges would largely be overcome following the discovery of the bacterial innate immune system of Streptococcus pyogenes and its adaptation for genome engineering (Hsu, Lander and Zhang 2014).

2. CRISPR history

The discovery of CRISPR began with the identification of a cluster of 29 bp repeats downstream of the iap gene locus in *E. coli* by Nakata and colleagues in 1987 (Ishino et al. 1987). These DNA sequences represented a unique form of clustered repeats found in >40% of bacterial species and later were collectively referred to as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) by Jansen and Mojica in 2002 (Jansen et al. 2002). Associated with these repeats were a number of CRISPR-associated genes (Cas) including Cas9, which was shown by Moineau and colleagues to be an RNA-directed endonuclease in 2010 (Garneau et al. 2010). The Cas9 protein of the Type II CRISPR system of *S. pyogene*, the most widely used for genome engineering, binds two RNAs encoded within the CRISPR repeats, the crRNA (CASCADE complex for type I; Cmr or Csm RAMP complexes for type III) and the tracrRNA (transactivating CRISPR RNA) (Deltcheva et al. 2011). The crRNA and the tracrRNA hybridize promoting cleavage by the host RNAsIII into a double-stranded hybrid crRNA/tracrRNA that when associated with the Cas9 protein acts to guide the cleavage of the target DNA 3 nucleotides 5′ to the
protospacer-adjacent motif (PAM); a sequence of 3 nucleotides adjacent to the binding site of the guide RNA (gRNA) within the target DNA, (e.g. NGG for S. pyogenes where N = any nucleotide). The crRNA/TracrRNA can be combined into a single chimeric guide RNA (gRNA) and Charpentier and Doudna demonstrated that this single gRNA could target site specific cleavage of DNA (Jinek et al. 2012). The CRISPR-Cas9 system was later applied to genome editing in mammalian cells by the groups of George Church and Feng Zhang in papers published in the same issue of Science in 2013 (Cong et al. 2013, Mali et al. 2013). Since the seminal work of the Charpentier/Doudna and Church/Zhang groups described above, the number of papers citing the use of CRISPR is doubling roughly every year with 1400 publications in 2015. The CRISPR-Cas9 system is extremely versatile and is being applied to not only gene knock-out studies, but to gene therapies such as muscular dystrophy (Long et al. 2014). As well enzymatically dead dCas9 can be used for gene regulation when it is fused to transcriptional activators and repressors (reviewed in (Dominguez, Lim and Qi 2016)) or epigenetic regulation through dCas9 targeting of histone modifying enzymes (Hilton et al. 2015), as well as chromosome tagging studies (Chen et al. 2013). In this review, we will discuss the advancements made in using SpCas9, its variants and related guided endonucleases to conduct precision genome editing.

3. Pathway choice in DNA repair affects the outcome of Cas9-mediated DNA breaks

The endonuclease activity of Cas9 enzymes create DNA DSBs that are recognized and repaired by cellular DNA repair machinery. In mammalian cells, DNA DSBs can be repaired by error prone NHEJ or by HR. During HR, a homologous DNA repair template, such as a sister chromatid or exogenous DNA, can be used for homology directed repair (HDR) and precise gene editing (Fig 1). However, there are many factors that affect the repair pathway choice after induction of a DNA DSB (Ceccaldi, Rondinelli and D'Andrea 2016) that can alter whether NHEJ or HR is favored, and this has been a very active area of research that is highly relevant in regard to enhancing rates of HDR during precision gene editing using CRISPR/Cas9. Therefore, to understand how to use CRISPR effectively to edit genomes, one must first understand the machinery and mechanisms underlying NHEJ and HR.
3.1. Canonical non-homologous end joining (c-NHEJ)

In mammalian cells, the predominate means of repairing DNA DSBs is by NHEJ (Chiruvella, Liang and Wilson 2013), which is in contrast to other organisms like yeast that are relatively poor at ligating blunt DNA ends and generally employ homologous sequences to repair breaks (Lieber 2010). The canonical NHEJ (c-NHEJ) repair pathway is activated in response to DNA breaks by the binding of the heterodimer Ku70/Ku80 to DNA ends and the DNA protein kinase catalytic subunit (DNA-PKcs), which together help to both protect the broken DNA ends and maintain them in close proximity to prevent translocations (Soutoglou et al. 2007, Roth, Lindahl and Gellert 1995) (Fig 1). DNA-PKcs can phosphorylate the histone variant H2AX within chromatin at the DSB to facilitate DNA lesion detection. However, the extent of H2AX phosphorylation during c-NHEJ appears to be more limited and may play a minor role in sensing and repairing DNA DSBs compared to other repair mechanisms such as alternative-NHEJ and HR (discussed below) (Lieber 2010). In contrast, autophosphorylation of DNA-PKcs and its transphosphorylation by the ataxia telangiectasia mutated (ATM) kinase are critical in c-NHEJ by facilitating the recruitment of Artemis, DNA pol lambda and mu to the repair site, and the regulation of Artemis 5’ exonuclease activity and cleavage of 3’ ssDNA overhangs to produce ligatable blunt dsDNA ends (Goodarzi et al. 2006, Jiang et al. 2015). Due to the activity of DNA pols and nucleases, DNA bases are added and removed during NHEJ, resulting in small insertions and deletions (indels) relative to the original genomic template (Fig 1). X-ray repair cross complementing protein 4 (XRCC4), XRCC4-like factor (XLF) and DNA ligase 4 (LIG4) are then recruited to the repair site where XLF helps LIG4 ligate the two DNA ends and complete the repair process (Chiruvella, Liang and Wilson 2013). Thus, while c-NHEJ is efficient and occurs throughout the cell cycle, it is error prone.

3.2 Alternative non-homologous end joining (alt-NHEJ)

In addition to c-NHEJ, the alternative NHEJ (alt-NHEJ) pathway, also known as microhomology-mediated end joining (MMEJ), is independent of KU70/80 and is initiated by larger stretches of
microhomology (4-14nt) (Lee-Theilen et al. 2011) (Fig 1). Alt-NHEJ uses different effector molecules and results in more dramatic errors post-repair such as larger deletions and possible gene rearrangements, including translocations between chromosomes (Zhang and Jasin 2011, Simsek et al. 2011). For example, during alt-NHEJ the poly ADP ribose polymerase 1 (PARP1) participates in the repair process, modifying proteins by ADP-ribosylation of proteins such as the nucleosome exchange factor SUPT16H (Heo et al. 2008) and by directly competing with Ku/DNA-PKcs for the binding of DNA ends (Wang et al. 2006, Heo et al. 2008). In fact, Ku70/80 can be repressive to alt-NHEJ and HR (Fukushima et al. 2001, Bennardo et al. 2008) and therefore competition between PARP1 and Ku/DNA-PKcs may influence which repair pathway is favored. In addition, PARP1 may enhance the retention and spread of H2AX on chromatin by modifying SUPT16H, thereby inhibiting the exchange of phosphorylated H2AX (γ-H2AX) in nucleosomes (Lieber 2010). This inhibition facilitates the recruitment of alternative DNA damage sensors such as the MRE11/RAD50/NBN (MRN) complex and thus shifts DNA repair pathway choice away from c-NHEJ (Lieber 2010). Unlike c-NHEJ, which can occur throughout the cell cycle, alt-NHEJ may be favored in the S/G2 cell cycle phases (Truong et al. 2013) similar to the HR pathway (discussed below) (Branzei and Foiani 2008, Mladenov et al. 2016). Alt-NHEJ also resembles HR in that both pathways involve limited end resection by the MRN complex and the nuclease CtIP (Zhang and Jasin 2011, Bennardo et al. 2008). End resection facilitates the exposure of DNA microhomology that contributes to the larger deletions, rearrangements and chromosomal translocations that occur during alt-NHEJ (Zhang and Jasin 2011, Bennardo et al. 2008). In further contrast to c-NHEJ, DNA ligase activity during Alt-NHEJ relies on DNA ligase 1 (LIG1) (microhomology-independent Alt-NHEJ) and DNA ligase 3 (LIG3) (microhomology-dependent Alt-NHEJ) rather than LIG4 (Simsek et al. 2011).

3.3 Homologous recombination

HR as discussed above is restricted to S/G2 of the cell cycle when sister chromatids are available as homology templates for the repair of DNA DSBs (Branzei and Foiani 2008, Mladenov et al. 2016). Despite the temporal restriction of HR during the cell cycle, phosphorylation of the histone variant H2AX
at the site of a DNA DSB is shared in common between both HR and NHEJ. H2AX is phosphorylated by the kinases ATM, ATM and Rad3-related (ATR) and DNA-PK (Durocher and Jackson 2001). γ-H2AX is then recognized by both the MRN complex (likely through NBS1 binding of γ-H2AX and/or via interaction with RAD17) (Kobayashi et al. 2002, Wang et al. 2014) and MDC1 (Stewart et al. 2003). However, there is evidence that the MRN complex can bind to DNA ends in the absence of γ-H2AX (Yuan and Chen 2010), and thus likely represents the initial recognition of the DNA break and subsequent coordination of the broken DNA ends in close proximity preventing chromosomal rearrangements (Soutoglou et al. 2007). If the DNA ends are damaged and unable to be easily ligated, the 5’ ends of the damaged DNA are resected by endonucleases such as MRE11/CtIP (short tract resection) or Exo1/BLM (long tract resection) to yield 3’ ssDNA (Mladenov et al. 2016) (Fig 1). These 3’ ssDNA ends are initially bound and stabilized by RPA but later replaced with RAD51, a process that requires the breast cancer 1 and 2 (BRCA1 and BRCA2) and the PALB2 proteins complexing with DNA (reviewed in (Savage and Harkin 2015, Prakash et al. 2015)). RAD51 forms DNA filaments on the 3’ ssDNA that enable homology search and invasion of homologous DNA to initiate the recombination process. Strand invasion of the homologous DNA template results in the formation of displacement loops (D-loops), DNA synthesis of homologous DNA by DNA polymerase delta and the formation of Holliday junctions. Resolution of the Holliday junctions requires nicking of the DNA backbones at the junctions and DNA ligation resulting in crossover (most common) and non-crossover (less common) resolution products (Reviewed in (Branzei and Foiani 2008)) (Fig 1).

3.4 Repair pathway choice

Although many DNA repair factors are cell cycle regulated, the molecular mechanisms that restrict HR to the S/G2 phases of the cell cycle and that determine repair pathway choice between NHEJ and HR are complex, and only partly elucidated (reviewed in (Ceccaldi, Rondinelli and D’Andrea 2016). Furthermore, competition between Ku70/80 and MRN for binding to DNA ends contributes to the regulation of repair pathway choice by controlling the ability of DNA ends to undergo end resection by
MRN/CtIP (Ceccaldi, Rondinelli and D’Andrea 2016, Chiruvella, Liang and Wilson 2013). For example, CtIP is controlled by CDK phosphorylation, restricting resection activity needed for HDR to the S/G2 phases (Huertas and Jackson 2009, Chen et al. 2008). DNA-PKcs also regulates the choice between NHEJ and HDR by acting as an antagonist of DNA end processing (Cui et al. 2005). However, various autophosphorylation states of DNA-PKcs can alter its association with DNA ends and can promote either NHEJ or HDR (Cui et al. 2005, Uematsu et al. 2007, Zhou and Paull 2013). Similarly, 53BP1 and BRCA1 appear to work antagonistically for binding to DNA DSBs (Panier and Boulton 2014) with 53BP1 binding being favoured during G0/G1 where it inhibits HR. Although the mechanism for HR inhibition is not fully elucidated 53BP1, via its interaction with RIF1, can antagonize BRCA1 and CtIP recruitment to DNA breaks (EscribanobDiaz et al. 2013). In contrast, during S/G2, BRCA1 expression peaks (Ruffner and Verma 1997), as does its association with DNA DSBs, excluding 53BP1 and resulting in promotion of HDR (Chapman et al. 2012, Bunting et al. 2010). In addition, BRCA2 promotes RAD51 association with ssDNA (Shivji et al. 2009), an association that is also restricted to S/G2 phases of the cell cycle by regulation of the interaction between BRCA1 and the PALB-BRCA2 complex (Orthwein et al. 2015).

An understanding of the mechanistic and regulatory details of these various DSB repair pathways is a critical consideration in the choice of gene-editing strategy. The availability of the robust NHEJ repair pathway in mammalian cells, which operates throughout the cell cycle, has been best exploited for the introduction of gene disrupting indels in target cells as an effective way of disrupting or “knocking-out” genes in cell and animal models. The precision of HDR however, is better suited to “knock-in” strategies and site-specific gene correction or mutation, however the low efficiency of HDR and its cell cycle dependence has proven a barrier for more ambitious gene therapy strategies.

4. Cas9-mediated gene editing using NHEJ

4.1 NHEJ-mediated Knock-outs
While non-homologous end joining is not the most precise method for genome engineering, it has the advantage over HDR-dependent methods in that it is much more efficient, can work in non-cycling cells and does not require a homology repair template, which can be laborious to generate. Using the more streamlined Cas9-mediated method to generate gene knock-outs by taking advantage of the error-prone nature of NHEJ, has therefore become a very popular approach for disrupting a gene of interest as an alternative to shRNA and other gene silencing techniques. As well, this approach is scalable and amenable to genome-wide knock-down screens (Hart et al. 2015, Gilbert et al. 2014, Shalem et al. 2014). The basic principle for gene knock-out is to use the NHEJ pathway to create insertions or deletions around specific parts of target genes (i.e. near start codons, common exons) which will result in loss of gene expression either through decreased transcription, loss of translation initiation, inappropriate mRNA production or truncated protein products (Fig 2A). Guide RNA selection is key to the success of NHEJ strategies with consideration given to the placement of the guide RNA and limiting off-target cutting (reviewed in (Graham and Root 2015)). For gene knock-outs, when possible, guide RNA’s should be placed wholly within a target exon, within the first half of the coding region of the protein, in a known functional domain, or near the ATG start codon (Doench et al. 2014, Shi et al. 2015). Although the NGG PAM motif of SpCas9 should be found about every 8 nt (Jinek et al. 2012), the G-C bias of the PAM sequence limits the availability of potential cut sites in A-T rich regions of the genome. However there are many Cas9 and CRISPR variants to choose from with different PAM requirements (detailed in section 5) that greatly increases the number of potential genomic cut sites. To simplify the gRNA design process, many of these design parameters have been incorporated into the over 20 online resources available for the identification and assessment of gRNAs for SpCas9 and other Cas9 variants (reviewed in (Graham and Root 2015, Bolukbasi, Gupta and Wolfe 2016)). Many of these resources such as CHOPCHOP (Labun et al. 2016), CRISPR design tool (Hsu et al. 2013), and E-Crisp Design (Heigwer, Kerr and Boutros 2014), will identify suitable gRNAs for knock-out experiments as well as predict and identify potential off-target sites. However, it remains a challenge to predict with complete accuracy how well a gRNA will recruit Cas9 to on-target or off-target loci or the extent of an indel within a target gene.
(Doench et al. 2016). Therefore, for gene knock-outs using NHEJ, multiple gRNAs should be evaluated and at least two should be used to generate independent knock-out cells. The issue of off-target effects may become less of an issue with the increased use and validation of high-fidelity Cas9 variants (Slaymaker et al. 2016, Kleinstiver et al. 2016) but the nature of the indels within a target gene will remain difficult to predict and thus will require validation and characterization. Fortunately there are many assays available to assess on-target and off-target efficiencies including the Surveyor- and T7E1-based assays as well as genome sequencing (Yu et al. 2014, Pilato et al. 2012). Furthermore, even if indels can be confirmed at the target locus, it is important to further analyze the edited genomic sequence for potential frameshifts, truncations, alternate AUG codons and/or splicing alterations in addition to characterizing the cell line for loss of transcript (e.g. by RT-PCR) and protein expression (e.g. by Western blotting).

4.2 NHEJ-mediated deletions, inversions and translocations

Gene knock-outs can also be generated by creating larger genomic deletions. Two different guide RNAs can direct Cas9 to two different loci on the same chromosome, resulting in two DSBs and large-scale deletions (several Kb up to 30 Mb) when the DNA ends are repaired through the NHEJ pathway (Cong et al. 2013, Gostissa et al. 2014, Canver et al. 2014, Fujii et al. 2013, Essletzbichler et al. 2014, He et al. 2015, Wang et al. 2015b). DNA ends on the same chromosome are repaired by NHEJ, resulting in deletion of the DNA sequences that were flanked by the breakpoints. This approach can be used to delete whole genomic regions or single exons of a gene target, resulting in gene disruption (Fig 2B). For example, a 105 Kb region of the tyrosinase (TYR) gene was deleted in rabbits (Song et al. 2016b), and a 65 Kb region containing the Dip2a gene was deleted in mice (Zhang et al. 2015).

Alternatively, using Cas9 and multiple gRNAs to create inter- and intra-chromosomal DSBs can be used to promote chromosomal translocations and gene inversions that can be used to model specific disease or cancer phenotypes (Renouf et al. 2014) (Fig. 2). For example, this strategy has been used in mice to create the Eml4-Alk gene inversion associated with non-small cell lung carcinomas (Blasco et al.
2014, Maddalo et al. 2014) and to model the human alveolar rhabdomyosarcoma translocation creating a Pax3-Foxo1 fusion protein (Lagutina et al. 2015). Similarly, Cas9- and NHEJ-mediated chromosomal translocations have been used to model Ewing’s sarcoma (EWSR1–FLI1 fusion gene) (Fig 2C), acute lymphoblastic leukemia (RUNX1-ETO fusion gene), and several lung cancer associated gene rearrangements (CD74-ROS1, EML4-ALK, KIF5B-RET) in human cells (Torres et al. 2014, Choi and Meyerson 2014).

4.3 NHEJ-mediated Knock-ins

NHEJ can also be exploited to produce kilobase-sized genomic insertions of a donor DNA sequence in a homology-independent manner. In the presence of a linear donor DNA sequence (e.g. linearized plasmid DNA), Cas9-generated DSBs can be ligated to the exposed ends of the linear DNA donor. This homology-independent NHEJ-dependent mechanism was first employed to insert transgenes into zebrafish and xenopus and allows for efficient insertion of larger DNA fragments than are often obtainable with HDR-based repair methods (Auer et al. 2014). Recently, He et al. used the same approach to generate very large (>10Kb) insertions in human cell lines and embryonic stem cells (He et al. 2016). In these systems, Cas9 is used to cut both the genomic DNA at the desired insertion site as well as the donor DNA plasmid (via distinct gRNAs) which facilitates the integration and ligation of the donor DNA at the genomic break point. This integration can occur in either orientation and will contain indels at the junctions between donor and genomic DNA (Fig 2D). Inserts as large as 34Kb are possible and this method is also effective at targeting silenced loci (He et al. 2016). The lack of control over copy number and orientation of the insert, and the presence of indels, prevents this from being a precise genome engineering strategy, however this strategy would allow for the introduction of complete gene expression or reporter systems (e.g. PuroR expression cassettes, DNA repair reporter assays). Targeting such systems to a safe-harbour locus such as adeno-associated virus integration site (AAVS1), located between exons 1 and 2 of the PPP1R12C gene on chromosome 19, would avoid the random integration concerns with retro- or lentiviral delivery based systems (Kotin, Linden and Berns 1992, DeKelver et al. 2010).
5. Precision Genome Editing with Cas9 using Homology Directed Repair (HDR)

While NHEJ is more efficient and better suited to large-scale knock-out studies, it lacks the precision required for more sophisticated genome engineering. The high fidelity of the HR pathway can be exploited to allow the insertion, deletion and substitution of a single nucleotide or large tracks of genomic DNA by HDR. As a result, HDR-based gene editing methods hold the greatest promise for the development of safe and highly precise gene therapy approaches to human disease.

A basic strategy for introducing gene editing using Cas9 and HDR requires three components: a Cas9 or variant to produce a double strand break, a guide RNA to direct Cas9 to the target genomic region and a homology repair template containing the desired edited sequences. With these three elements, researchers can create controlled deletions or insertions to knock out genes or specific isoforms or engineer nucleotide changes to introduce or correct disease phenotypes such as β-thalassemia (Niu et al. 2016, Xu et al. 2015) and Alzheimer’s disease (Paquet et al. 2016). The HDR editing strategy can also introduce in-frame epitope tags for affinity purification (e.g. TAP tag), immunoprecipitation and immunodetection (e.g. FLAG, V5) or live cell microscopy (e.g. GFP). As well, promoters can be manipulated to remove or introduce transcription factor binding sites or place a gene under control of an inducible promoter (e.g. by introduction of tetracycline response elements (TREs) (Gossen and Bujard 1992). Genes can also be disrupted by the insertion of a genetic “payload” such as a cDNA encoding a GFP or puromycin resistance expression cassette. In addition, safe harbour loci, such as the AAVS1 locus, can be targeted for the insertion of complete gene loci, including promoters, cDNAs and polyadenylation signals. Insertion of complete genes allows for selection of target cells (e.g. by antibiotic resistance or fluorescence), as well as provides a means to validate gene knock-outs and explore protein structure/function by “add back” of wild-type or mutant target gene expression.

The choice of which combination of resources to use will depend on the desired change to the target genome as well as user preference; however, a common goal should be to efficiently and precisely edit a target locus while minimizing off-target events. In this section we provide an overview of the elements
required to conduct Cas9-HDR-based genome editing and summarize some of experimental tools and
approaches that have been used.

5.1 Choosing an enzyme:

Since the original characterization of Cas9 nucleases for genome editing were published in 2013,
many more Cas9 and related nucleases have been described and made available to the research
community, much of which is driven by the goal of decreased off-target editing. The first modifications to
humanized *S. pyogenes* Cas9 (hSpCas9) to minimize off target cutting were the D10A and H840A
mutations which each eliminate the activity of one of the two nuclease domains and thus convert Cas9
into a nickase (Cas9n) capable of cutting only one strand of the DNA backbone (Ran et al. 2013b, Jinek et
al. 2012, Cong et al. 2013). These variants are used with paired guide RNAs that will direct the Cas9n
enzymes to nick the DNA backbone on opposite strands in close proximity (-4 to 20 nt offset) which will
produce a double strand break (Ran et al. 2013b). Off target effects are reduced 50-1500-fold (Ran et al.
2013b) since the probability of the targeting gRNAs directing two Cas9n in similar close proximity
elsewhere in the genome is greatly reduced. A similar approach to the nCas9 strategy is to conjugate
enzymatically dead Cas9 (dCas9-D10A, H840A) to one half of the homodimeric FokI endonuclease. In
this approach, two gRNAs are used to direct the dCas9-FokI fusion protein (fCas9) to target DNA loci
(about 30nt apart) where the FokI nuclease can only produce a DNA DSB if it dimerizes, thus eliminating
the potential for unwanted single strand breaks generated by Cas9n (Guilinger, Thompson and Liu2014,
Tsai et al. 2014). While the nickase approach greatly reduces unwanted off-target cutting, the requirement
to have two gRNA’s with PAM motifs on opposite strands with a short off-set sequence can sometimes
be too restrictive and place undesirable limits on the genome editing design strategy.

More recently, two high fidelity variants of SpCas9 have been described (Slaymaker et al. 2016,
Kleinstiver et al. 2016). The enhanced specificity eSp-Cas9(1.1) was engineered with three point
mutations (K848A, K1003A and R1060A) that remove positive charges that are likely involved in
stabilizing the interaction between the non-target DNA strand and Cas9 (Slaymaker et al. 2016). By
weakening in the ability of Cas9 to bind the non-target strand, it could favor displacement of the gRNA on the target strand and re-annealing of the dsDNA when there are mismatches between the gRNA and the target DNA strand. eSpCas9(1.1) retains on-target activity vs. WT Cas9 but virtually abolishes off-target effects. Another recently described high fidelity Cas9 (SpCas9-HF1) was designed by mutating four amino acid residues that interact with the DNA backbone (N497, R661, Q695, Q926) to alanine, thus weakening Cas9-DNA interactions and reducing the tolerance for off-target binding (Kleinstiver et al. 2016). These mutations were also shown to synergize with the D1135E mutation of SpCas9, which enhances specificity for the NGG PAM motif to further reduce off-target cutting (Kleinstiver et al. 2016, Kleinstiver et al. 2015). Similar to eSpCas9(1.1), SpCas9-HF1 retains on-target activity while greatly reducing off-target activity. Since the high fidelity Cas9’s show similar or superior reduction in off-target effects as the nickase variants and require only a single gRNA, it is likely the use of high fidelity enzymes will become the gold-standard approach for gene editing.

There are several other Cas9 variants isolated from various bacterial species available to researchers for genome editing as well. Although these variants are not as well characterized as SpCas9, many show similar nuclease activity and, more importantly, many recognize non-NGG PAM motifs. For example, Cas9 from Staphylococcus aureus (Sa) recognizes NGRRT or NGRRN (Ran et al. 2015), Neisseria meningitidis (Nm) recognizes NNNNGATT (Zhang et al. 2013), Streptococcus thermophilus (St) recognizes NNAGAAW (Horvath et al. 2008, Deveau et al. 2008) and Treponema denticola (Td) recognizes NAAAAC (Esvelt et al. 2013). As well, SpCas9 has been engineered to recognize different PAM motifs such as the VQR (NGAN or NGNG), EQR (NGAG) and VRER (NGCG) variants (Kleinstiver et al. 2015). Similarly, an alternative Class II type V-A cas enzyme called Cpf1 has recently been characterized (Zetsche et al. 2015). Cpf1 is isolated from Francisella novicida and differs from SpCas9 in that the PAM motif is T-rich (5’-TTN-3’) and the nuclease cuts the DNA in a staggered fashion creating a 5 nucleotide 5’ overhang starting 18 nucleotides 3’ of the PAM. The ability of Cpf1 to generate DNA overhangs that could act as ‘sticky ends’ has been speculated to potentially offer advantages for HR-based gene editing. However, no data supports this assumption and it is likely that
Cpf1-induced breaks will be subject to both HR and NHEJ mechanisms in a similar fashion to SceI-induced DNA breaks. However, it may be possible to design strategies to take advantage of the DNA overhangs generated by Cpf1 to allow NHEJ-based directional insertion of DNA at specific genetic loci.

5.2 Guide RNA design considerations

Basic Cas9 gRNA design is relatively simple with the minimum requirements being a 20nt sequence upstream (5’) of an NGG protospacer adjacent motif (PAM) (Deltcheva et al. 2011, Jinek et al. 2012). The PAM is essential for Cas9 activity while specificity is derived from the 20nt targeting sequence. Cas9 enzymatic activity is directed to cut the DNA backbone between the 3rd and 4th nucleotide upstream of the PAM motif (Jinek et al. 2012). There is some tolerance for mismatches which can contribute to off target effects and increased fidelity has been observed using shorter (17-18mer) sequences (Fu et al. 2014). With the introduction of many academic and commercial online resources the design of gRNA sequences is becoming much faster and simpler (Graham and Root 2015). Guide RNA design resources such as those discussed in section 4, are useful for designing and choosing appropriate gRNAs with minimal predicted off-target cutting for HDR-based genome editing. However, for HDR-based editing, particularly when using nickases, the location of potential gRNAs can be restricted to small regions (50-100nt) around a desired cut site. For efficient HDR it is desirable to cut as close to the location of the desired change as possible (Elliott et al. 1998). Therefore, while online resources are useful for predicting off-target effects, compromises must sometimes be made to ensure efficient on-target editing. Consideration can also be given to introducing mutations to the PAM and gRNA target site in the homology repair template to prevent recutting once editing has taken place. If these mutations occur in coding regions, conservative mutations that avoid changing the coding sequence should be engineered into the homology arms. When using Cas9n nickase strategies, the need for generation of the DSB near the junction of the two homology arms provides a fortuitous opportunity to engineer the homology donor plasmid such that each gRNA target sequence is positioned in a different homology arm.
and separated by the payload sequence. This prevents Cas9n from cutting the donor vector and the correctly edited genome since the targeted DNA nicks will be too far apart to generate a DSB.

5.3 Designing Homology repair templates

The most critical element of HDR mediated genome editing is the homology repair template that will contain the desired changes to the target genome. A basic homology repair template must contain sufficient homology to the regions both upstream and downstream of the genomic breakpoint, which can be referred to as the 5’ and 3’ homology arms. Mutations and insertions can be engineered by placing altered sequences between these homology arms. Deletions can be obtained by designing homology arms that flank the desired deleted sequence in the genome. The two main types of homology repair templates are single stranded oligodeoxynucleotides (ssODN), used for making small genetic changes (about 50-100nt) and homology donor plasmids which are used for making larger changes up to several kilobases.

When working with cell lines it is advisable to first sequence the region around where the editing will occur, to account for cell-specific mutations and single nucleotide polymorphisms (SNPs) that can reduce targeting efficiency due to the effects of mismatches (homeology) between the target gene locus and donor template (Tham, Kanaar and Lebbink 2016). Even one mismatch in 100 bases is sufficient to reduce HR by up to 6-fold (Elliott et al. 1998). One approach to mitigate problems with DNA mismatches it to amplify homology arms from the genomic DNA of the target cell or synthesize them based on sequence analysis of the target locus from the same cell line to be edited.

5.3.1 ssODNs for point mutations and small changes: (<100bp) - Small genomic edits such as point mutations or introduction of small tags (e.g. HA, FLAG) can be introduced using ssODN as the homology repair template (Fig 2E). These ssODNs are typically less than 200nt long making them affordable and easy to synthesize. The repair template consists of short (30-60 nt each) homology arms flanking the small insert or point mutation (1-100 nt) (Yang et al. 2013, Ran et al. 2013a). It is important to design the homology repair template such that the ends of genomic DSB is as close to the region of homology as
possible since the efficiency of gene conversion decreases dramatically with distance from the break (Elliott et al. 1998). These distances should be less than 10nt but larger distances up to 100 nt can work, but with reduced efficiency (Elliott et al. 1998). Recent insight into the kinetics and mechanism of Cas9-mediated DNA cleavage has revealed that Cas9 can locally release the 3’ end of the non-target DNA strand, making it accessible to ssODN donor sequences that are complementary to the non-target strand (Richardson et al. 2016). As well, designing ssODNs with homology asymmetrically (91 nt PAM-proximal, 36nt PAM-distal) placed around the genomic breakpoint can facilitate excellent HDR rates of up to 60% (Richardson et al. 2016). Since ssODNs can be easily synthesized, gene editing strategies that use them to target and tag genes in their endogenous loci (e.g. HA or GFP11 tags) are amenable to large scale or high-throughput applications (Mikuni et al. 2016, Leonetti et al. 2016).

5.3.2 Homology Repair Template Donor Plasmids for Large Genetic Changes – The ability to use Cas9 to precisely engineer and insert, or “knock-in” large DNA fragments into a specific target genomic locus has opened up many exciting new possibilities for building research tools and developing novel gene therapy strategies. The homology repair templates required for these strategies usually consists of a homology donor plasmid containing a specific genetic payload (i.e. insertion) flanked by two homology arms. The primary homology template design considerations include determining how large to make the homology arms, designing the genetic insert to produce the desired effect in edited cells and developing a strategy for assembling these elements into a donor plasmid.

In general, increasing homology arm length increases the efficiency of homologous recombination (Hasty, Rivera-Perez and Bradley 1991, Thomas and Capecchi 1987, Shulman, Nissen and Collins 1990). Similarly, larger homology arm lengths, up to about 2 Kb, support more efficient knock-in of genetic payloads (Wang et al. 2015a, Li et al. 2014). For insert sizes of about 1-2 kb, homology arms of 500-800nt are sufficient and, as a general rule, having each homology arm about 50-100% the size of the payload (up to 2 Kb per homology arm) appears to provide suitable HDR results (Pinder, Salsman and Dellaire 2015, Li et al. 2014, Ratz et al. 2015, Mikuni et al. 2016). For smaller insertions (<1 Kb), smaller
homology arms of about 200-500nt can be easily synthesized commercially, whereas larger homology arms should be amplified from genomic DNA for the cell line of interest. Insert or payload sizes of about 1-2 Kb are frequently reported and this is generally large enough to add a protein tag (e.g. GFP, 720 nt) or a simple gene expression cassette with a promoter and poly adenylation signal (Li et al. 2014, Pinder, Salsman and Dellaire 2015, Ratz et al. 2015). However, several groups have successfully knocked-in larger genetic payloads of between 3 and 8 Kb using homology donor plasmids which approaches the 10 Kb insert sizes obtainable using NHEJ knock-ins (Wang et al. 2015a, Lee et al. 2016, Zhang et al. 2015).

In the following subsections, we will discuss some of the specific applications of using HDR to correct Cas9-mediated DSBs to conduct genomic editing, which are also summarized in Figure 2.

**Protein Tagging**: The nature and the purpose of the donor sequence knock-in can vary dramatically based on the gene engineering strategy (Fig 2F). Perhaps one of the more practical uses of Cas9-mediated knock-in strategies is the development of improved research tools and cell line model systems. The ability to introduce peptide and protein tags on a gene of interest that can be expressed at endogenous levels provides a mechanism for avoiding the pitfalls of over-expression based studies (Dambournet et al. 2014, Ratz et al. 2015). For example, the coding sequence for reversibly switchable EGFP (rsEGFP2) was inserted in-frame with the high mobility group protein HMG-I (HMGA1), Vimentin (VIM), Zyxin (ZYX) genes to create C-terminal GFP-tagged fusion proteins expressed from their endogenous promoters (Ratz et al. 2015). Similarly, we tagged the PML protein with the GFP-variant clover by inserting the clover sequence in frame after the second codon of exon 1, which is common to all PML isoforms (Pinder, Salsman and Dellaire 2015). Since all isoforms are tagged and expressed from the endogenous promoter, this system in an improvement over previous methods involving overexpression of a single PML isoform (Pinder, Salsman and Dellaire 2015). In addition, using Cas9 and HDR to tag a protein with an affinity tag (e.g. TAP) in the endogenous locus can facilitate the identification of protein complexes by mass spec under more physiological conditions (Dalvai et al. 2015). For these strategies, the gene knock-in consists of the coding sequence of the protein tag that must be inserted in-frame in the coding sequence of the
target gene. The homology arms, therefore are designed to immediately flank the insertion site with no loss of genomic information (Dambournet et al. 2014).

**Gene Disruption:** As an alternative to using NHEJ to create gene-disrupting indels, Cas9-mediated DNA DSBs can be repaired by HDR to introduce specific gene disrupting reporter genes. For example, a puromycin resistance gene or fluorescent protein gene can be inserted in-frame into a coding exon of a target gene (Fig 2G). If this sequence is followed by a stop codon, one can terminate expression of the target gene and use the inserted sequence to select for correctly edited cells. This strategy would use the endogenous promoter, and so would be most useful on genes that are highly and constitutively expressed. As with protein-tagging strategies, the homology arms and gene knock-in sequence must be in-frame. Alternatively, one could insert an entire gene expression cassette, complete with a promoter and/or polyadenylation signal (Hammond et al. 2016). This could be targeted to a common exon, however, preservation of reading frame with the targeted gene is not necessary. These expression cassette containing donor constructs tend to be larger (>2Kb) and since the vector can drive its own expression, it is possible that selected cells may harbour a non-specific random integration of the homology donor. Thus, when possible gene disruption by “gene-trapping” of the target gene using a promoterless homology donor is preferable. Another caution with this strategy is the observation that larger insertions into exons can result in exon skipping, potentially leading to altered splicing of the target protein, rather than gene disruption (Uddin et al. 2015).

**Safe Harbour Loci:** In addition to the AAVS1 locus in the PPP1R12C gene on chromosome 19, other genomic safe harbour loci include the CCR5 and ROSA26 loci on chromosome 3 in humans and the Rosa26 locus on chromosome 6 in mice (reviewed in (Sadelain, Papapetrou and Bushman 2011). These genomic loci provide a better characterized and controlled site for the integration of various genetic payloads, than can be provided with the random integration of virus-based gene delivery vectors. As well, larger DNA payloads can be inserted using Cas9-HDR knock-in strategies than with retroviral (~8kb) or
Adeno-associated virus (~5Kb) delivery systems. These loci, particularly AAVS1, can be used with increasingly standardized tools such as gRNAs targeting the locus, as well as homology repair donor templates and plasmids with well-defined and characterized homology arms. Although the deliverable genetic payloads may vary, the genomic locus and the tools to target it can be kept constant between experiments, allowing for more predictable and comparable results. The AAVS1 locus can be used to insert a variety of gene expression cassettes (e.g. cDNA), reporter assays (e.g. HDR, Luciferase) or selection markers (puromycin resistance, GFP) (Dalvai et al. 2015, Oceguera-Yanez et al. 2016, Castano et al. 2016, Zhu, Gonzalez and Huangfu 2014) (Fig 2H). It should be noted that although AAVS1, CCR5 and ROSA26 have been used successfully for the construction of improved research tools, these safe-harbour loci have not been extensively evaluated for their appropriateness for more therapeutic applications where disruption of the genes at or near the target loci may have adverse effects (Papapetrou and Schambach 2016).

**Large deletions**: Large deletions can be introduced using HDR as well as NHEJ. Similar to the NHEJ strategies discussed in section 5, generating deletions using Cas9 and HDR requires engineering two DNA DSBs flanking the region to be deleted. Whereas the NHEJ method will introduce indels at the ligation site, HDR could be used to create a precise junction between the ends of the excised DNA (Fig 2I) or introduce a new DNA sequence to replace the excised DNA fragment. This method could be used to delete specific gene features (e.g. promoter elements, exons) or to delete entire genes. For example, Zhang et al. used Cas9 to generate two DSBs at the murine Dip2a gene locus resulting in excision of a 65 Kb fragment that was repaired by HDR using a donor plasmid that inserted a neomycin resistance cassette between the exposed genomic ends (Zhang et al. 2015).

The above examples only cover some of the potential applications of using HDR for genome engineering strategies. These approaches are most practical for editing the genomes of cell lines to build more powerful and precise research tools. However, these methods have also been applied to human stem
cell populations (Li et al. 2015, Schwank et al. 2013, Xie et al. 2014) and can therefore be used to develop
gene therapies and disease models as well.

6. Further Optimization Strategies for HDR

The greatest limitation on the use of HDR in genome editing is the low efficiency of HDR compared
to NHEJ. This is in part due to the fact that HDR occurs primarily in the S/G2 phases of the cell cycle, a
limitation that also restricts genome editing efforts in non-proliferating cells. The main strategies for
overcoming this limitation includes inhibiting NHEJ, enhancing HDR efficiency, and, as discussed in the
previous section, introducing selectable makers into editing strategies to enrich for cells that have
undergone the rare HDR events.

6.1 NHEJ inhibitors and HDR enhancers

A particularly attractive strategy for shifting the NHEJ-HDR balance in favor of HDR by
genetically or pharmacologically suppressing NHEJ or enhancing HR (Fig 1, red and green arrows). This
concept was initially employed with the earlier genome editing systems of ZNFs and TALENs in plants
and insects, where the mutation or RNAi-mediated knock-down of KU or LIG4 gene expression was used
to impair NHEJ and thus enhance HDR efficiency (Qi et al. 2013, Basu et al. 2015). Later
pharmacological inhibition of LIG4 by the small molecule SCR7 was used to suppress NHEJ in
mammalian cells in conjunction with the Cas9 system, resulting in ~2 to ~20-fold enhancement of HDR
and genome editing (Chu et al. 2015, Robert et al. 2015, Maruyama et al. 2015). However, we (Pinder,
Salsman and Dellaire 2015) and others (Song et al. 2016a, Gutschner et al. 2016, Lee et al. 2016, Yang et
al. 2016a) have found limited enhancement of Cas9-mediated HDR with SCR7 suggesting that this drug
is not universally useful for enhancing HDR in all experimental systems. Nonetheless, NHEJ suppression
by various methods has been demonstrated by several groups to enhance Cas9-mediated HDR. For
example, NHEJ suppression has also been accomplished through inhibition of DNA-PKcs using the small
molecules NU7441 (Leahy et al. 2004) and Ku-0060648 (Munck et al. 2012) resulting in ~2-4 fold
enhancement of Cas9-mediated HDR in HEK293T cells and mouse embryonic fibroblasts (MEFs) using donor plasmids or ssODNs as repair templates (Robert et al. 2015). The small molecules Brefeldin A, an ER-Golgi transport inhibitor (Ktistakis, Linder and Roth 1992) and L755507, a β3-adrenergic receptor agonist (Parmee et al. 1998), also enhance Cas9-mediated HDR ~2 to 9-fold through uncharacterized mechanisms (Yu et al. 2015). We have recently shown that enhancement of HDR using the Rad51-stimulating small molecule RS1 (Jayathilaka et al. 2008) can increase Cas9-mediated HDR ~3-6-fold in HEK293 and U2OS cells (Pinder, Salsman and Dellaire 2015). RS1 is also effective at enhancing Cas9-mediated HDR ~2-5-fold in rabbit embryos (Song et al. 2016a). Using small molecules to enhance HDR provides a technically simple way to improve Cas9-mediated precision genome editing. The above studies generally report little to no cell toxicity at effective doses for these small molecules. However, the suppression of NHEJ or enhancement of HDR by these compounds could result in accumulation of other undesired mutations resulting from off-target breaks or random DSBs generated as a normal part of cell division.

Alternatively, since many genome editing systems are transfection based, it is convenient to use this same method to further modify NHEJ and HDR in target cells by introducing siRNA, shRNA or expression plasmids targeting these repair pathways. For example, Chu et al. used shRNA to silence Ku70, Ku80, and Lig4, alone and in combination, resulting in suppression of NHEJ and a 2-5-fold enhancement of HDR (Chu et al. 2015). Similarly, Robert et al. used siRNA to silence Lig4, DNA-PK, Ku70 and Ku80 resulting in suppression of NHEJ and an approximately 2-fold increase in HDR (Robert et al. 2015). Another mechanism for suppression of NHEJ is expression of the adenovirus serotype 4 (Ad4) E1B55K and E4orf6 proteins which promote the proteosomal degradation of Lig4 (Forrester et al. 2011, Cheng et al. 2011, Schwartz et al. 2008). Co-expression of Ad4 (Chu et al. 2015) or Ad5 (Robert et al. 2015) E4orf6 and E1B55K enhanced HDR by about 7-fold and 3.5-fold respectively. Interestingly, E1B55K and E4orf6 from the Ad5 and other serotypes can also degrade p53 and Mre11 (Cheng et al. 2011). Mre11 loss could reduce HDR efficiency by limiting MRN binding to DSBs. The effects of p53 loss are more difficult to predict. In response to DNA damage, p53 can induce cell cycle arrest, which
could positively or negatively impact HDR efficiency depending on which cell cycle checkpoint is
affected by p53 loss (i.e. G1/S, G2/M; see section 6.2 below) (Speidel 2015). HDR can also be
suppressed by p53, therefore p53 loss could enhance HDR (Menon and Povirk 2014, Sirbu et al. 2011,
Willers et al. 2000). Although both groups report reduced NHEJ and increased HDR with expression of
the Ad E1B55K and E4orf6 proteins, a side-by-side comparison of Ad4 and Ad5 E1B55K and E4orf6
coexpression in the same experimental system would be potentially enlightening as to shared
mechanisms. However, the multiple cellular targets for these viral proteins make them less desirable as an
approach to enhancing HDR in more therapeutic or tightly controlled genome editing applications. Using
a similar over-expression approach, we attempted to promote HDR by expressing BRCA1 variants that
exhibit a hyper-recombination phenotype (Pinder, Salsman and Dellaire 2015, Dever et al. 2011,
Escribano-Diaz et al. 2013). BRCA1 is important for influencing DNA repair pathway choice and
expression of wild-type BRCA1 or the K1702M or M1774R hyper-recombination variants had a marginal
positive effect on HDR in our system with the M1775R variant showing ~3-fold increase in HDR (Pinder,
Salsman and Dellaire 2015). Finally, overexpression of Rad51 can enhance HR (Yu et al. 2011) and thus,
expression of Rad51 alone, or in combination with the small molecule RS1 may further enhance HDR of
Cas9-mediated DSBs.

6.2 Cell Cycle Considerations

Another strategy for enhancing HDR is to promote and prolong the S and G2 phases in target
cells when HDR is most active. There are several well-established methods for synchronizing cells and
preventing progression through the various stages of cell cycle and these have been applied to Cas9-
genome editing strategies. The Doudna lab (Lin et al. 2014) screened six chemical cell cycle inhibitors for
the ability to improve HDR-mediated repair of Cas9-induced DSBs. HEK293T cells were synchronized
with Nocodazole (G2/M), lovastatin (M/G1), mimosine (G1/S), aphidocholin (G1/S), thymidine (G1/S)
and hydroxyurea (G1/S.) prior to nucleofection 24hrs later with Cas9-gRNA ribonucleoprotein (RNP)
complexes targeting the EMX1 locus and a ssODN repair template. There was a 2- to 4-fold enhancement
of HDR with Nocodazole synchronization (Lin et al. 2014). Similar enhancement of HDR (3- to 6-fold) was achieved with the microtubule polymerization inhibitors Nocodazole or ABT-751 (Yoshimatsu et al. 1997, Yang et al. 2016a). These drugs were used to synchronize human pluripotent stem cells and neural precursor cells prior to nucleofection with Cas9/gRNA expression plasmids and homology repair templates (Yang et al. 2016a).

Another approach for enhancing HDR is to minimize Cas9 activity during G1 phase when HDR is suppressed and NHEJ prevails. Gutschner et al. (Gutschner et al. 2016) engineered a Cas9 variant that is fused to the fragment of human Geminin (Cas9-Gem) that is recognized by the APC/Cdh1 E3-ligase complex during G1, targeting Cas9 for degradation by the proteasome. During S/G2/M, the APC/Cdh1 complex is inactivated and Cas9 becomes stabilized, thus promoting the generation of DNA DSBs during the cell cycle phase when HDR is most active. An 87% enhancement of HDR was observed with the Cas9-Gem over WT-Cas9 using an EGFP reporter system at the AAVS1 locus in HEK-293T cells (Gutschner et al. 2016) While this increase was modest, it was cumulative with other cell cycle interventions, such as Nocodazole treatment, indicating that these combinatorial approaches may be necessary to achieve maximal increase in HDR efficiency. Notably, Nocodazole was used in this study to arrest cells at the G2/M phase 31hrs after transfection of Cas9, gRNA and donor plasmids, rather than to synchronize cells prior to delivery of Cas9 RNPs as described above; however, the increase in HDR efficiency was more modest at about 1.5-fold (Gutschner et al. 2016).

For non-cycling cells a recent study (Orthwein et al. 2015) described a novel mechanisms of HR inactivation by the KEAP1/CUL3 ubiquitin E3 ligase complex in G1 and provide a mechanism through which HDR can be reactivated in non-cycling cells. In G0/G1 cells, BRCA1 recruitment to DNA breaks is inhibited by 53BP1 (Escribano-Diaz et al. 2013), CtIP is not active as the appropriate CDKs responsible for its phosphorylation are not active (Huertas and Jackson 2009, Chen et al. 2008), and BRCA1’s interaction with the BRCA2:PALB2 complex (required for RAD51 filament formation) is inhibited by ubiquitination of PALB2 by a KEAP1/CUL3 complex (Orthwein et al. 2015). Thus, to overcome inhibition of HDR in G0/G1 cells, Orthwein and colleagues overexpressed activated CtIP
endonuclease (T847E mutant) with concomitant depletion by siRNA of the DNA repair protein 53BP1 and the KEAP1/CUL3 complex, which together reversed HR repression and increased HDR in G1 cells (Orthwein et al. 2015). These observations provide the basis for new strategies for editing the genomes of non-cycling cells such as neurons or muscle cells. In addition, this same approach could help reactivate HDR in the G1 phase of cycling cells, further enhancing Cas9-mediated HDR.

The above strategies to enhance HDR have proven successful yielding increases in HDR in the order of 2-20-fold. For some experimental systems, this enhancement should be sufficient for making selection of correctly edited clones significantly less laborious with overall HDR rates in some cell populations reaching 20-60% (Chu et al. 2015, Maruyama et al. 2015, Lin et al. 2014). However, combining various HDR strategies has generally shown only modest increases in HDR enhancement to 5-10-fold overall (Pinder, Salsman and Dellaire 2015, Chu et al. 2015, Robert et al. 2015). This might suggest that there is a practical experimental ceiling to how much HDR can be improved in each experimental system which could be influenced by cell/model-specific recombination activities, cell cycle profile and efficiency of cell transfection. As described in section 5, the introduction of a selectable marker, which can be targeted to the AAVS1 safe harbour locus, can aid in the selection of correctly edited cells. Ideally, however, it would be better to not have to include these potentially confounding variables in a gene editing strategy, particularly when engineering cells for therapeutic, rather than basic research purposes. Thus, finding better ways to promote HDR remains an important area of genome editing research.

7. Recent Discoveries and Innovations

Interest and research in genome-editing has rapidly expanded with the discovery of bacterial-derived programmable nucleases like Cas9 and with the wide distribution of gene-editing reagents in plasmid repositories such as Addgene (Kamens 2015). Below we have chosen to highlight two recent advances in
gene-editing technologies that illustrate the creativity of the research in this field and the great promise of even more advances in genome-editing in the near future.

7.1 dCas9-directed Base Editing

A novel approach to making point mutations comes from Komor et al (Komor et al. 2016) using nuclease dead cas9 (dCas9 D10A H840A) fused to rat APOBEC1 cytidine deaminase. This base-editing strategy does not rely on homology directed repair or require a repair template, but rather relies on the enzymatic activity of APOBEC1 to generate C-U mutations in the exposed ssDNA non-target strand that is displaced when the Cas9-gRNA complex forms R-loops (Fig 2J). This results in C-T or G-A mutations following DNA replication. One potential drawback of this strategy is that the deaminase is processive and can convert all C’s to U’s within an ~5 base window near the distal end of the protospacer from the PAM. Thus, editing may occur on undesired cytosine bases nearby.

7.2 Beyond Cas9: Bacterial DNA-guided endonucleases

A new addition to the oligo-nucleotide guided endonuclease tool-box has been reported by Gao et al (Gao et al. 2016a) who describe the use of the Argonaute protein from Natronobacterium gregoryi (NgAgo) as a DNA-directed endonuclease capable of facilitating genome editing in eukaryotic cells. NgAgo differs from Cas9 in a number of ways that make NgAgo an exciting addition to the genome editing toolbox. First, NgAgo uses a single stranded 5’ phosphorylated DNA molecule of about 24 nucleotides as a guide, rather than the single stranded RNAs required by Cas9. Using a DNA guide, rather than an RNA guide, appears to have the advantage for increasing cleavage of G-C-rich sequences because the DNA guide is less prone to formation of inhibitory secondary structures than RNA. In addition, NgAgo DNA guides do not require a PAM motif or have other target sequence requirements, which removes one of the primary constraints found with directing Cas9 activity to specific genomic loci. NgAgo is also not very tolerant of mismatches between the guide DNA and the target sequence, with a single mismatch reducing activity by >70% and three mismatches completely ablating DNA cutting (Gao
et al. 2016a). In order for the guide DNA to become associated with NgAgo, it must be present at the time of NgAgo synthesis, since pre-formed NgAgo does not associate with suitable guide DNA molecules in cells or in vitro. Another notable feature of NgAgo is the removal of several nucleotides from the target region during the DNA cleavage process. This feature may enhance gene knock-out strategies relying on NHEJ since indels are desirable however, is it not likely to adversely affect HDR-based repair strategies since the deleted nucleotides can be re-introduced with the homology repair template. Deletion of sequences from the target site is likely to prevent re-cutting as NgAgo is intolerant of mismatches.

8. Concluding Thoughts

In the short time since the initial publications describing the use of Cas9 for editing human genomes in 2013, many challenges have already been met and overcome. Initially CRISPR research focused on discovering novel Cas9 enzymes and defining the PAM motifs that are essential for directing the Cas9 and related enzymes to their target DNA sequences. This was followed by X-ray crystallographic structures of Cas9 in complex with guide RNA and DNA targets (Nishimasu et al. 2014, Anders et al. 2014) which helped to define the critical amino acids responsible for targeting and would support the rational design of mutant Cas9 enzymes with altered PAM specificity, and increased fidelity. In particular, the use of high fidelity Cas9 variants with dramatically reduced off-target effects are predicted to rapidly become the “gold-standard”, and, along with robust web-based gRNA design tools, Cas9-mediated targeted gene disruption is approaching the ease of use of RNA interference for reverse genetic studies.

One outcome of this rapidly evolving field, has been the use of Cas9-mediated HDR by researchers to build increasingly powerful and sophisticated cell line and animal research tools in virtually any DNA-based organism. However, improving HDR efficiency and targeting gene editing machinery in specific cell and tissue populations in vivo remains one of the major challenges to the genome editing field and a main barrier to safe and effective human gene therapy. For now, the most successful gene therapies in humans will involve diseases and disorders of the hematopoietic system, where the host system can be
completely ablated and reconstituted from edited hematopoietic stem cell populations generated and validated *in vitro*. This could involve correcting gene mutations involved in blood diseases such as sickle cell anemia (Huang et al. 2015) and β-thalassemia (Xie et al. 2014, Yang et al. 2016b), or the engineering of HIV-resistant hematopoietic stem cells (Kang et al. 2015). Similarly, engineering of embryonic stem (ES) cells or patient-derived induced pluripotent stem (iPS) cells could be used to help treat symptoms or preserve function in patients with muscle and nervous system disorders, an approach already being developed for Duchenne muscular dystrophy using CRSIRP/Cas9 (Li et al. 2015) and myotonic dystrophy using TALENs (Gao et al. 2016b). Finally, genome editing can also be used to improve immunotherapies for cancer. For example, patient T-cells can be engineered to carry chimeric antigen receptors (CARs) targeting leukemias and other cancers. This gene therapy, known as CAR-T cell therapy, is already in clinical trials for B-Cell leukemia employing viruses to integrate an anti-CD19 CAR that targets B-Cells randomly into the genome of patient T-Cells (Park, Geyer and Brentjens 2016).

However, the random nature of viral integration poses some safety risks, including unintended cell mutations that could cause secondary cancers in treated patients. Thus, a safer approach to CAR-T cell therapy may be the use of precision genome editing by Cas9-mediated HDR to generate engineered T-Cells without the use of viral integration. However, to facilitate Cas9-mediated gene therapy, increasingly sophisticated gene and viral delivery strategies may be needed to fully realize the therapeutic potential of this technology, and thus represent a very important area of future study.
Figure 1.
Figure 1. DNA repair pathway choice for correcting Cas9-mediated DSBs. Cas9 and related nuclease can be directed to specific genomic loci through association with a guide RNA molecule. Generation of double strand breaks (DSBs) results in exposed DNA ends that become bound by the end-recognition protein complexes consisting of Ku70/80 or MRN (MRE11/RAD50/NBS1). When bound by Ku70/80, repair proceeds through classical non-homologous end joining (c-NHEJ), which is the major DNA repair pathway, particularly in G1 cells. Recruitment of DNA-PKcs and Artemis and DNA polymerase (Polλ/μ) coordinates end processing resulting in small insertions and deletions of less than 20 nucleotides. The ends are then ligated with the DNA ligase 4 (LIG4), XRCC1 and XLF1 complex. Suppression of NHEJ using siRNA, small molecules (SCR7, NU7441, Ku-0060648) or adenoviral proteins E4orf6 and E1B55K (shown in red) can result in increased DNA repair though alternative NHEJ (alt-NHEJ) or homologous recombination (HR). DNA repair using the alt-NHEJ pathway begins with the recognition of DNA ends by the MRN complex. Recruitment of PARP1 and CtIP results in limited 5’ end resection and exposure of 3’ single stranded DNA (ssDNA) overhangs that can locally interact with regions of microhomology (4-14 nucleotides). XRCC1 and DNA ligase 3 (Lig3) contribute to the resolution of the repair focus, resulting in end ligation with larger genomic deletions compared to c-NHEJ). During the S and G2 phases of the cell cycle, when sister chromatids are present, DNA repair can proceed through homologous recombination. Recognition of DSBs by the MRN complex initiates the HR process where recruitment of BRCA1 and CtIP facilitate limited 5’ end resection (short tract). This is followed by longer end resection by the BLM/EXO1 protein complex (long tract). The exposed ssDNA is initially bound and stabilized by the RPA protein which is then exchanged for Rad51 with the assistance of BRCA1, BRCA2 and PALB2. Rad51-bound ssDNA can then search and invade homologous DNA such as sister chromatid or foreign homology repair template such as plasmid DNA. Resolution of the homologous recombination process results in faithful incorporation of the homologous sequence with the broken DNA ends. This process can be exploited to introduce novel DNA sequences into the genomic DNA at the site of the DSB. Homologous recombination can be enhanced through the use of small molecules such as brefeldin A, L755507 and the Rad51 activator RS1 or over expression of BRCA1 and other HR components (shown in green).
Figure 2.
Figure 2. Exploiting NHEJ and HR for precision genome editing with Cas9. (A) Effective gene knock-out strategies use guide RNA molecules that direct Cas9 to the coding region of a gene of interest, resulting in DNA breaks that are repaired by the error-prone NHEJ pathway. The resulting indels can cause various mutations (truncation, frame-shift, etc) that disrupt protein expression or function. (B) Larger deletions up to 1Mb can be generated by targeting Cas9 to two points on the same DNA strand, resulting in loss of the intervening genomic fragment when the distal ends are ligated together. (C) Chromosomal translocations can also be engineered in a similar way, by directing Cas9 to make DSBs on two different chromosomes. (D) In addition, large insertions of foreign DNA can be introduced by providing linearized double-stranded DNA fragments in the presence of Cas9-mediated DSBs. The ends of the linear DNA fragment can become inserted in either orientation and ligated to the exposed chromosomal DNA ends and ligated via NHEJ. In the presence of a homology repair template, Cas9-mediated DSBs can be repaired through homologous directed repair (HDR). (E) Smaller genetic changes (<100 nt) can be introduced using single stranded oligodeoxynucleotides (ssODNs) to repair Cas9-mediated DSBs. (F) Repair templates can be engineered to facilitate the in-frame knock-in of novel genetic sequences such as fluorescent protein tags (e.g. GFP). (G) Gene disruptions can be generated by inserting an in-frame selectable marker (e.g. puromycin resistance (PuroR)) followed by a termination codon into a coding exon of a gene of interest, thereby simultaneously preventing the target gene from being correctly translated and providing a mechanism to select for cells harbouring the gene disruption. (H) Genetic safe-harbour loci, such as AAVS1, provide a predictable environment for the introduction of complete gene expression cassettes and reporter systems. (I) Large genomic deletions can also be engineered by generating two distant DNA DSBs and designing a repair template with no insert to join the distal DNA strands. (J) Finally, single base editing can be accomplished using catalytically dead Cas9 fused to cytidine deaminase to generate C to T mutations.

Acknowledgements

This research was funded by operating grants from the Canadian Institutes of Health Research (CIHR) to G.D. (MOP-84260). GD is a Senior Scientist of the Beatrice Hunter Cancer Research Institute (BHCRI). J.S. was a recipient of a BHCRI postdoctoral fellowship made possible by the Harvey Graham Cancer Research Fund as part of the Terry Fox Foundation Strategic Health Research Training Program in Cancer Research at CIHR.
REFERENCES:


Bunting, S.F., Callen, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N.,
Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C.X., Finkel, T., Nussenzweig, M., Stark, J.M. and
Nussenzweig, A. 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking

Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H. and Orkin,
S.H. 2014. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced
palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 289(31): 21312-
21324. doi: 10.1074/jbc.M114.564625 [doi].


Castano, J., Herrero, A.B., Bursen, A., Gonzalez, F., Marschalek, R., Gutierrez, N.C. and Menendez, P.
2016. Expression of MLL-AF4 or AF4-MLL fusions does not impact the efficiency of DNA damage

Ceccaldi, R., Rondinelli, B. and D'Andrea, A.D. 2016. Repair Pathway Choices and Consequences at the

Chapman, J.R., Sossick, A.J., Boulton, S.J. and Jackson, S.P. 2012. BRCA1-associated exclusion of
53BP1 from DNA damage sites underlies temporal control of DNA repair. J. Cell. Sci. 125(Pt 15): 3529-
3534. doi: jcs.105353 [pii].

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H.,
Weissman, J.S., Qi, L.S. and Huang, B. 2013. Dynamic imaging of genomic loci in living human cells by


Panning, B., Ploegh, H.L., Bassik, M.C., Qi, L.S., Kampmann, M. and Weissman, J.S. 2014. Genome-
Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 159(3): 647-661. doi:
10.1016/j.cell.2014.09.029 [doi].

N., Jeggo, P.A. and Lees-Miller, S.P. 2006. DNA-PK autophosphorylation facilitates Artemis
endonuclease activity. EMBO J. 25(16): 3880-3889. doi: 7601255 [pii].

Gossen, M. and Bujard, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-

Gostissa, M., Schwer, B., Chang, A., Dong, J., Meyers, R.M., Marecki, G.T., Choi, V.W., Chiarle, R.,
Zarrin, A.A. and Alt, F.W. 2014. IgH class switching exploits a general property of two DNA breaks to
10.1073/pnas.1324176111 [doi].


nuclease improves the specificity of genome modification. Nat. Biotechnol. 32(6): 577-582. doi:
10.1038/nbt.2909 [doi].

10.1016/j.celrep.2016.01.019 [doi].


transgene expression in human iPSCs and their differentiated derivatives. Methods. 101: 43-55. doi:
10.1016/j.ymeth.2015.12.012 [doi].

Orthwein, A., Noordermeer, S.M., Wilson, M.D., Landry, S., Enchev, R.I., Sherker, A., Munro, M.,
Pinder, J., Salsman, J., Dellaire, G., Xia, B., Peter, M. and Durocher, D. 2015. A mechanism for the
suppression of homologous recombination in G1 cells. Nature. 528(7582): 422-426. doi:
10.1038/nature16142 [doi].

Cell Biol. 15(1): 7-18. doi: 10.1038/nrm3719 [doi].


Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S. and
Tessier-Lavigne, M. 2016. Efficient introduction of specific homozygous and heterozygous mutations

Park, J.H., Geyer, M.B. and Brentjens, R.J. 2016. CD19-targeted CAR T-cell therapeutics for
hematologic malignancies: interpreting clinical outcomes to date. Blood. doi: blood-2016-02-629063
[pii].


Pilato, B., De Summa, S., Danza, K., Papadimitriou, S., Zaccagna, P., Paradiso, A. and Tommasi, S.
2012. DHPLC/SURVEYOR nuclease: a sensitive, rapid and affordable method to analyze BRCA1 and

https://mc06.manuscriptcentral.com/bcb-pubs


Schwank, G., Koo, B.K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S.,
repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell. Stem

82(18): 9043-9055. doi: 10.1128/JVI.00925-08 [doi].

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L.,
cells. Science. 343(6166): 84-87. doi: 10.1126/science.1247005 [doi].

10.1038/nbt.3235 [doi].

Shivji, M.K., Mukund, S.R., Rajendra, E., Chen, S., Short, J.M., Savill, J., Kleenerman, D. and
Venkitaraman, A.R. 2009. The BRC repeats of human BRCA2 differentially regulate RAD51 binding on
106(32): 13254-13259. doi: 10.1073/pnas.0906208106 [doi].

Shulman, M.J., Nissen, L. and Collins, C. 1990. Homologous recombination in hybridoma cells:

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar,
E.J., Gregory, P.D., Holmes, M.C. and Jasin, M. 2011. DNA ligase III promotes alternative


