Published in final edited form as:

Bioessays. 2019 December ; 41(12): e1900126. doi:10.1002/bies.201900126.

Can Designer Indels Be Tailored by Gene Editing?:

Can Indels Be Customized?

Sara G. Trimidal,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Ronald Benjamin,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Ji Eun Bae,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Mira V. Han,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Elizabeth Kong,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Aaron Singer,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Tyler S. Williams,

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

martin.schiller@unlv.edu.

[[†]]Present address: Nevada Institute of Personalized Medicine and School of Life Sciences University of Nevada Las Vegas, 4505 S. Maryland Pkwy, Las Vegas, NV 89154-4004, USA

Conflict of Interest

The authors declare no conflict of interest.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/bies.201900126>

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Bing Yang,

Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

Martin R. Schiller^[+]

School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Abstract

Genome editing with engineered nucleases (GEENs) introduce site-specific DNA double-strand breaks (DSBs) and repairs DSBs via nonhomologous end-joining (NHEJ) pathways that eventually create indels (insertions/deletions) in a genome. Whether the features of indels resulting from gene editing could be customized is asked. A review of the literature reveals how gene editing technologies via NHEJ pathways impact gene editing. The survey consolidates a body of literature that suggests that the type (insertion, deletion, and complex) and the approximate length of indel edits can be somewhat customized with different GEENs and by manipulating the expression of key NHEJ genes. Structural data suggest that binding of GEENs to DNA may interfere with binding of key components of DNA repair complexes, favoring either classical- or alternative-NHEJ. The hypotheses have some limitations, but if validated, will enable scientists to better control indel makeup, holding promise for basic science and clinical applications of gene editing. Also see the video abstract here <https://youtu.be/vTkJtUsLi3w>

Keywords

CRISPR/Cas9; double strand break; gene editing; meganuclease; nonhomologous end joining; TALEN; zinc finger nuclease

1. Introduction

Genome editing with engineered nucleases (GEENs) is a biotechnology for manipulating genomes in cells and organisms. GEEN technologies have several different types of applications and many review articles detail advancements enabled by GEENs.^[1–6] For example, genome editing broadens species range for generation of engineered mutant organisms. Knockout or knockin animals with an exogenous gene or a marker inserted at a specific locus can be readily created with gene editing. Gene editing also has therapeutic potential for the removal of toxic alleles or insertion of functional copies to rescue nonfunctional alleles. Furthermore, viral infections can be targeted by gene editing.^[7–10]

All GEEN technologies are based upon targeting a protein or protein/RNA (ribonucleoprotein) complex to recognize a specific DNA sequence, endonucleolytic cleavage producing one or more DNA double-strand breaks (DSB)s, DSB repair, and eventually site-specific genetic alterations in a genome.^[1,11] GEEN technologies are mainly based on four types of customizable nucleases: 1) meganucleases; 2) zinc finger nucleases

(ZFNs); 3) transcription activator-like effector nucleases (TALENs); and 4) RNA-guided clustered regularly interspaced short palindromic repeats (CRISPR) associated endonucleases (e.g., Cas9, Cas12a, and so on; Figure 1).

Meganucleases are natural or engineered homing endonucleases that recognize DNA binding sites of 12–40 base pairs (bp), longer than that of a typical restriction enzyme (Figure 1).^[12,13] Zinc finger nucleases (ZFNs) are chimera of three to six zinc finger repeats in which each repeat binds a specific DNA base triplet, and they are fused to the nonspecific DNA cleavage domain of the FokI endonuclease.^[14]

Over the years, the FokI catalytic domain, specificity, and affinity have been improved through directed protein evolution.^[15] TALENs are similar to ZFNs with chimera of 12–24 transcription activator-like effector (TALE) repeats with each repeat binding to a single DNA base, fused to the DNA cleavage domain of FokI.^[16–18]

CRISPR/Cas9 is a more recently developed genome editing technology comprising a guide RNA bound by Cas9 or another RNA-guided endonuclease, and is derived from the bacterial immune system.^[3,19] The guide RNA (gRNA) determines the DNA binding specificity and the Cas9 complex cleaves the DNA. The part of the gRNA complementary to the target DNA sequence is known as Protospacer. Additionally, Cas9-mediated DNA cleavage requires a Cas9 binding signal known as Protospacer Adjacent motif (PAM sequence), that recognizes the 2–5 bp long PAM sequence located downstream of the desired DNA target sequence. Differences in the specificity of targeting, flexibility in sequence recognition, off-target editing, implementation time, and delivery method are important factors to consider when choosing among GEEN technologies.^[1,3,20,21]

GEEN-induced DSBs are repaired in eukaryotic cells by two main DNA repair systems: 1) accurate homology directed repair (HDR) and 2) more error-prone repair via nonhomologous end joining (NHEJ).^[22] In genome editing, indels (insertions and deletions) are introduced upon repair by NHEJ.^[23] There are two different types of NHEJ mechanisms or pathways that produce indels: 1) canonical NHEJ (C-NHEJ); and 2) alternative NHEJ (A-NHEJ, B-NHEJ, or A-EJ). A-NHEJ can occasionally mediate microhomology-mediated end-joining (MMEJ), a form that requires microhomology (A-NHEJ-MMEJ).^[24] Variable (diversity) joining (V(D)J) recombination is a specialized form of C-NHEJ active during somatic recombination of the genome in B-cell immunoglobulins and T-cell receptors.^[22,25,26] V(D)J recombination is not discussed further hereafter, because it is only active in immune cells.

1.1. What Are the Types of NHEJ Repair Pathways?

The C-NHEJ repair pathway is the primary NHEJ pathway active in the repair of environmentally induced DSBs.^[27,28] The C-NHEJ pathway components are KU70, KU80, DNA-PK, Artemis, and Werner syndrome protein (WRN) nuclease, DNA polymerases μ and λ , Terminal Transferase (TdT), X-ray repair cross-complementing protein 4 (XRCC4), XLF, and Ligase IV (LIGIV). Although we later include Ataxia telangiectasia mutated (ATM) and Breast cancer type 1 susceptibility protein (BRCA1) with A-NHEJ, there is also support for potential dual roles in C-NHEJ (Table S1, Supporting Information).^[29–33] KU70 and KU80

form a protein complex recruited to DSBs and serves as a scaffold for recruitment of additional C-NHEJ factors and end ligation by the XRCC4/XLF/LIGIV complex. LIGIV is the only ligase known for the C-NHEJ pathway.^[34]

The A-NHEJ pathway comprises poly(ADP-ribose) polymerase 1 (PARP1), CTBP interacting protein (CTiP), Exonuclease 1 (Exo1), Ligase I (LIGI), Ligase III (LIGIII), the MRN complex consisting of MRE11, NBS1, and RAD50, and breast cancer type 1 susceptibility protein (BRCA1) (Table S1, Supporting Information). In A-NHEJ, the MRN complex recognizes DSBs, and then recruits ATM, ATR, CTiP, and BRCA1.^[35] End processing proceeds by recruited endonucleases CTiP and Exo1, as well as Pol β DNA polymerase. The XRCC1/LIGI or /LIGIII complexes ligate the processed DSB ends.

Some DSB repairs use microhomology with short stretches of complementary single-strand sequences (1–20 bp) annealing, similar to the extensive complementary DNA stretches that anneal in HDR. Basic A-NHEJ (discussed above) does not use microhomology (A-NHEJ), whereas A-NHEJ/MMEJ can (Table S1, Supporting Information).^[24] Genomic breakage cleaved by GEENs can be repaired in vivo by either C-NHEJ, A-NHEJ, and A-NHEJ/MMEJ.

1.2. Are Indel Composition and Features Random?

After GEEN-induced damage, each NHEJ pathway introduces new indels upon repair. The indels can vary in length from 1 bp to 10 kb, but short microindels of 1–50 bp are much more frequently observed (referred to as indels hereafter).^[36] The lengths of resulting deletion indels are dictated by NHEJ pathway-specific exonucleases that digest the ends of DSB, thus the deletion length is related to the extent of exonuclease digestion prior to ligation. The protein complex and exonuclease constituents of C-NHEJ and A-NHEJ differ, which may potentially affect the length of deletion indels. The length of insertion indels is dependent on the same exonucleases, but also dependent upon NHEJ pathway-specific DNA polymerases or terminal transferases, some of which are template-independent.^[37–40]

There are several potential benefits to control the lengths and types of NHEJ indel edits with different GEENs and their variable architectures. Better control over indels could advance the generation of model organisms for research. Designer indels with frameshifts of 1 or 2 bp insertion(s) or deletion(s) would increase the probability of generating gene knockouts and reduce the downstream effort of screening cells, clonal cell lines, and organisms. Designer indels of variable lengths could specifically target functional elements such as exons, introns, promoters, and terminators.

A better understanding of indel signatures can also guide interpretation of genome edits. Furthermore, understanding indels can increase knowledge about mechanisms and signatures of NHEJ repair, which would help scientists study transposons and NHEJ editing. Indels are a historical record and can provide insight into the editing history of genomes. Indel analysis could be used to infer the historical activity of different DNA repair enzymes in evolution. Several software tools can aid in predicting or interpreting indels analyzed with machine learning models such as Indelphi, FORECasT (favored outcomes of repair events at

Cas9 targets), and Menthumicrohomology predictor tool algorithm. For instance, Indelphi can predict 1–60 bp deletions and 1 bp insertions with high accuracy.^[41–46]

2. Indels Impact Disease

Indels can cause human disease and are important in cancer pathology.^[47] Indels represent ~18% of human genome variation.^[48,49] About 75% of indels are generated by polymerase slippage during DNA replication.^[50] Slippage produces tandem repeat expansion, homopolymer runs, and microsatellite instability.^[47,51] The remaining 25% of indels are thought to arise mostly from NHEJ.^[48]

In addition to NHEJ, a smaller percentage of non-repetitive indels are introduced by retrotransposition, the insertion of transposable elements after reverse transcription. These insertions can be up to 6 kb in the case of full-length Long Inter-spersed Nuclear Elements (LINEs), but are usually truncated to a few hundred bases. Host DNA repair pathways repair the 5' end of the insertion after retrotransposition. Knockdown of ATM, ERCC1/XPF, or other core proteins of C-NHEJ demonstrates necessity for retrotransposition.^[52–55] Analysis of sequences at the 5' junction of retrotransposition shows that there are three types of sequence features at the LINE integration site, 1) microhomology of 1–2 bp, 2) insertion of extra nucleotides, or 3) no extra sequence, indicating there are distinct pathways involved in the repair.^[56] Whether the pathways repairing retrotransposition are the same as the known C-NHEJ or A-NHEJ pathways will require further investigation. Occasionally, retrotransposition can generate deletions at the endonuclease cleavage site that range from a few bases up to 1 Mb.^[57]

While little is known about the control of NHEJ indel type, length, and nucleotide composition, indels with insertions of 1–3 bp are a known signature of C-NHEJ repair.^[58] In this review, we consider the evidence that could be used to modify cell systems to control indel types and length, as well as interpret non-repeat indel signatures. We do not cover the nucleotide composition of indels because the NHEJ polymerases do not have strongly differing affinities for different nucleotides and C-NHEJ has multiple polymerases and a terminal transferase. Thus, we expect it would be difficult to unambiguously control the nucleotide composition of indel insertions.

3. Do NHEJ Pathways Have Indel Signatures?

We reviewed the literature to determine whether manipulation of expression of NHEJ genes systematically alters the type and length of indels. A list of the studies and an analysis of Sanger sequencing of select clones is provided in File S1, Supporting Information.

3.1. Can We Select for Insertion or Deletion Indels?

Indels types can be short insertions or deletions, or complex indels having an insertion together with a deletion. Environmental damage by carbon ion beams, X-rays and gamma-ray irradiation causes DSBs that are repaired by NHEJ creating new indels (Figure 2). A review of the literature reveals that deletions resulting from environmentally induced damage are 10 times more frequent than insertions. Deletion () or knockdown (–) of many

different NHEJ genes drastically influences the preference for insertions or deletions. In C-NHEJ, reduced expression of KU80, Pol λ , or LigIV strongly favors deletions, whereas reduced expression of DNA-PK or ATM, and overexpression of Pol λ or TdT favors insertions (Figure 2). In A-NHEJ, reduced expression of the MRN complex (NBS1, RAD50, MRE11), as well as CTIP and overexpression of Exo1 favors insertion indels, while reduced expression of LigI, LigIII, and BRCA1 produces deletion indels. Although there are caveats, especially considering the different model systems studied and the generally small number of edits examined, these studies suggest that expression of many NHEJ genes can be manipulated to control the ratio of deletion to insertion edits.

3.2. Manipulating Indel Edit Lengths with C-NHEJ Genes

Indels with insertions of 1–3 bp are known signatures of C-NHEJ repair.^[58] Gene expression mutants for several C-NHEJ genes produce editing profiles that are dramatically different from control cells (Figure 3, and Table S2, Supporting Information). In cells lacking functional KU70, KU80, or LigIV, few insertion events are observed indicating that these proteins are required for insertions and that C-NHEJ is a principal pathway for genetic insertion lesions (Figure 3).^[40,59,60]

Specific polymerases or exonucleases are randomly recruited to DSBs and dictate insertions and deletions during C-NHEJ repair.^[28,40] C-NHEJ polymerases μ and λ add nucleotides in insertion edits and TdT adds 1–2 bp insertions.^[60,61] DNA Pol μ and Pol λ synthesize inserts in a template-independent manner.^[37–40] Pol λ cells completely lack insertions, indicating that Pol λ is required for all C-NHEJ insertions.^[62] In addition, Pol μ cells still have a low frequency of short insertions similar to controls.^[63] Cells overexpressing TdT (+TdT) have a much higher fraction of mid-sized insertions up to 9 bp, although TdT is only known to catalyze 1–2 bp insertions.^[60,64] Other C-NHEJ nulls and knock-downs do not have a dramatic effect on the editing profile, suggesting that they are not required for C-NHEJ mediated insertion edits.

As expected, many of the same genes for insertions in C-NHEJ are also required for deletions. *KU70* cells lacked deletion edits, but all other C-NHEJ genes had different types of deletions (Figure 2). Like *KU70*, the *KU80*, *DNA-PK*, *XRCC4*, *ATM*, and *Polλ* null mutants had few, or no short deletions (1–3 bp), confirming the necessity of C-NHEJ in these edits (Table S2, Supporting Information). However, 2–3 bp deletions are not blocked in *-LIGIV* cells, suggesting that A-NHEJ may be responsible for 2–3 bp deletions (Figure 3). Knockout of several C-NHEJ genes (*KU80*, *DNA-PK*, *ATM*, and *XRCC4*) increases the frequency of 5–11 bp deletion edits, suggesting that A-NHEJ is responsible for these mid-sized deletions. Deletions are due to exonuclease digestion of DSBs. In addition to Artemis, the WRN exonuclease excises nucleotides in C-NHEJ deletions, but no editing profile data was available for these genes.^[28,65,66] Knockout of *Polμ* and *Polλ* increases the frequency of very long deletion edits > 50 bp, likely favoring the use of exonucleases in end processing. A summary of possible C-NHEJ pathway signatures and mutants that support them are in Table S2 and Figure S7, Supporting Information.

3.3. Controlling Indel Type and Length through A-NHEJ/MMEJ

Knockdown of *LIGI*, *LIGIII*, *BRCA1*, or *MRE11* produces few or no insertions, confirming a necessary role for A-NHEJ, like C-NHEJ in all lengths of insertions (Figure 3C).^[30,67,68] Consistent with this conclusion, knockout of *Exo1* produces only insertions and no deletions. However, *NBS1*, *RAD50*, and *CTIP* knockdown cells have observable insertion edits, but these proteins are also necessary for HDR, which has cross-talk with NHEJ.^[69]

The A-NHEJ pathway is also required for some deletion edits. Specifically, knockdown of *NBS1*, *RAD50*, and *BRCA1* eliminates short 1–3 bp deletion edits (Figure 3D)—this is similar to the essential role of C-NHEJ in these short edits. –NBS1, –RAD50, –MRE11, –LIGI, and –LIGIII mutants reduce the frequency of these very long > 50 bp deletion edits supporting a A-NHEJ mechanism for such edits. The number of short 1–3 bp and > 50 bp deletions present in control conditions are absent in *Exo1* nulls consistent with its exonuclease activity. However, we note that –CTIP, and –BRCA1 are required for longer deletion edits (Figure 3D).

Collectively, these results implicate A-NHEJ in 1–3 bp deletions and long deletions >50 bp. An analysis of MMEJ did not lead to any significant conclusions (Figures S2–S5, Supporting Information).

3.4. Longer Microindels When Both C-NHEJ and A-NHEJ Are Inhibited

When key genes required for both the C-NHEJ (*XRCC4*) and A-NHEJ (*MRE11*) pathways are deleted or suppressed (called C-NHEJ/-A-NHEJ), NHEJ editing remains intact and the preference for deletion or insertion indels does not change (Figure 2).^[25,69,70] However, these cells in which both pathways are inhibited have insertions of 5–17 bp that are not present in control cells and a high prevalence of 8–13 bp deletion edits (Table S2 and Figures S1 and S7, Supporting Information).^[71]

Thus, in conclusion, reduction or elimination of key C-NHEJ and A-NHEJ genes might prove useful for producing longer deletion and insertion microindels. Although NHEJ gene manipulation cannot yet be used to fine-tune the length of indels, a body of published literature suggests that the lengths of both insertion and deletions edits can be modulated to a higher level of granularity. A summary of the findings that control indels is shown in Figure 4. Potential sequence signatures are summarized in Results and Figure S7, Supporting Information.

4. GEENs Likely Influence NHEJ Repair Mechanisms

Structural studies of NHEJ proteins in complex with DNA yield insight into NHEJ repair. The initial DSB recognizing protein complex, KU70/KU80 recruits other C-NHEJ pathway components. In a 3D structure, the KU70/KU80 complex occupies a 14 bp binding site on the DNA, which is displaced ~8 bp from the DSB break end (PDB accession no.: 1JEQ; Figure 5A). In A-NHEJ, recruitment of PARP1 to the lesion initiates repair.^[24] An X-ray crystal structure of the PARP1/DNA complex shows that PARP1 binds the terminal 4 bp of DNA (4AV1; Figure 5B). The WRN/DNA complexes show that the WRN helicase binds the terminal 2 bp stretch of DNA (3AAF). Analysis of the structural data and DNA binding

locations infer repair of DSBs after cleavage by some GEEN technologies may select specifically for C-NHEJ or A-NHEJ.

4.1. ZFN Editing May Reduce Shorter Indels

ZFNs may saturate and remain bound to the DNA binding sites after cleavage, which is likely because the binding affinities for ZFNs are in the low nanomolar range.^[72] If the ZFNs remain bound to the DSB, the ZFN would have a steric clash for loading the KU70/KU80 heterodimer, because the binding site for this protein would overlap with part of the ZFN binding site (Figure 5A). Therefore, ZFNs would compete with this complex for the same binding site, likely inhibiting the C-NHEJ pathway. Conversely, binding of A-NHEJ proteins to DSBs likely accommodates concurrent binding of ZFNs, consistent with spacers between the DNA binding and FokI endonuclease domains that are generally much longer than the 2–4 bp recognitions sites for PARP1 and WRN (Figure 5B). Thus, with the ZFN bound, there may be room to load the initial components of the A-NHEJ pathway. The A-NHEJ pathway may be preferentially selected in ZFN editing. A review of the literature indicates that most indels produced by ZFNs, occur in the ZFN binding sites, which would then enable subsequent end processing and ligation by either NHEJ pathway (Figure 5C). However, when compared to environmental damage, indels introduced by ZFNs had fewer 1–2 bp insertions and deletions, consistent with A-NHEJ repair (Figure 6A,B).

4.2 TALEN Editing May Prefer Short Insertions

Like ZFNs, TALENs also have a higher frequency of mid-sized 5–8 bp insertion edits (Figure 6B,C) and the same structural arguments apply as discussed above. However, repair of TALEN-induced damage prefers short 1–2 bp insertion edits. Since TALENs and ZFNs share the same FokI endonuclease producing the same type of DSB ends, any differences are likely in the DNA binding domains. The differences in editing profiles maybe because TALENs bind a much longer stretch of DNA than ZFNs or/and longer flexible linker requirements that connect FokI to TALE repeats.^[73] This may affect how other endogenous repair proteins are recruited to, and process DSBs.

4.3. Can Modified CRISPR/Cas9 and Meganucleases be Used to Generate Longer Indels?

The meganuclease binding site overlaps with that of KU80, as well as PARP and WRN (Figure 5A,B). Consistent with these structures, bound meganucleases may block access to NHEJ complexes. Nearly all published NHEJ indel edits resulting from meganucleases occur in the nuclease binding sites. This suggests that enzyme release opens access to NHEJ DSB recognition and repair machinery (Figure 5C). Several studies report that the enzyme remains bound to one of two DSB ends and has affinities in the nanomolar range. However, the other DSB end remains accessible to repair machinery, potentially explaining the observed edits in the binding site.^[74–78] Long deletion edits of > 25 bp were most evident in meganuclease-induced DSBs when compared to environmental damage or other GEENs (Figure 6). An increase in 5–8 bp insertion indels, such that with as ZFNs, was also observed.

Like meganucleases, the CRISPR/Cas9 DNA binding foot-print likely sterically interferes with binding of KU80 to DSBs in the C-NHEJ pathway, as well as PARP and WRN to

inhibit A-NHEJ (Figure 5A,B). CRISPR/Cas9 has a strong nanomolar binding affinity for its target DNA and stays bound to its cleavage products following digestion, at least in vitro (Figure 5).^[79–81] However, the CRISPR/Cas9 indel length profile is most similar to environmental damage, albeit having reduced frequencies of 1 bp deletions, (Figure 6A,E) and a slight preference for insertion indels (Figure 2). One explanation is that a small portion of DSBs that release CRISPR/Cas9 are available for NHEJ repair.

Some new Cas9 chimeras are also being used to control longer indels. A dimeric Cas9 chimera generates larger deletion indels of about ~200 bp with 97% efficiency.^[82] Fusion of the I-TevI to Cas9 (TevCas9) generates 33–36 bp deletions.^[83]

5. Does Recursive Editing Produce Complex Indels?

Another difference among GEEN technologies and between control is the number of complex indels. Complex indels contain a mixture of insertions and deletions in a single indel edit. Complex indels are determined by sequencing of clones isolated from a cell population. The mechanism of their formation remains uncertain, but clues are emerging. Environmental damage produces few complex indels (4.2%; Figure 5C), similar to cancer, a natural disease state.^[84] All GEENs produce a higher percentage of complex indels (7–11%). The percentage of complex indel edits for TALENs and ZFNs is similar to that previously reported.^[85] The vast majority of disease-related indels can be modeled by a two-step process with insertions, deletions, and substitutions.^[86]

Given these observations, a likely explanation for complex indels is repetitive editing events at the same site, enhanced by GEEN technologies that may recursively cleave the same site until new indels destroy their binding sites, as previously suggested.^[83,87] GEENs have a high affinity for DNA and some remain bound to DNA after cleavage, at least in vitro.^[77,81] However, many GEENs have high indel frequencies in the DNA binding site (Figure 5C) necessitating GEEN dissociation for repair. One possibility is that a new indel edit does not alter the GEEN binding site. Given the high affinity and high cellular concentrations of the GEEN, it is likely to rebind the site with the new indel and possibly introduce a new overlapping indel.^[88] This new indel may or may not impinge into the DNA binding site. This recursive mechanism likely proceeds until the GEEN binding site on the DNA is damaged and the GEEN cannot re-bind. This may be why almost all CRISPR/Cas9, and meganuclease edits are in the DNA binding site, but not necessarily in TALENs and ZFNs, where the binding sites are engaged by the TALE or ZF, respectively, and the cut site engaged by FokI. This hypothesis is supported by the observation that persistent editing with CRISPR/Cas9 yields longer microindels than basal NHEJ.^[89] This mechanism would also explain how both A-NHEJ and C-NHEJ can participate in generating the same complex indel. In fact, this mechanism could be extended to both insertion and deletion indels being edited by both NHEJ pathways.

6. Example Limitations in Analysis of Indels

Caveats in the analysis of gene editing and NHEJ presented herein should be considered with caution, and interesting findings should be verified through further experimental

testing. There are general limitations to our analyses such as publication bias and heterogeneity of results, which are inherent to other similar analysis.^[90] We would have preferred to reach conclusions based on a statistical analysis; however, this analysis was not possible because of the small and heterogeneous sample sizes of many supporting studies analyzing a small set of clones by Sanger sequencing likely had sampling bias, especially for the analysis of NHEJ pathway mutants presented herein. Another pitfall of Sanger sequencing is its low sensitivity to detect indels of larger size by PCR amplification.^[91]

Nevertheless, our approach was similar to that of Kim et al. that found similar differences in repair events between ZFN and TALEN gene edits.^[85] Future studies that utilize gene editing should analyze recent NGS data for NHEJ and HDR.

We pooled data for many different types of organisms and some species have differences in NHEJ. For example, human cells have substantially higher DNA-PK activity than rodent cells.^[92,93] However, the majority of the data we analyzed are for vertebrates, which generally have similar pathways.^[94] Therefore, we did not observe any significant species-specific difference in editing (Figure S6, Supporting Information).

GEENs cleave DNA producing DSB termini that are blunt, sticky, or sticky with microhomology, and these ends may be differentially processed, even by the same NHEJ pathway. The architecture of recombinant GEEN proteins themselves may differ in a way that effects editing profiles. For example, TALENs and ZFNs can be designed to have different length linkers between the DNA binding domain and nuclease that impact editing.^[95] GEEN delivery and expression level can also affect repair.^[89]

7. Conclusions and Outlook

The main conclusions are:

- a. Indels are a major source of genetic variations.
- b. The indel type (insertion, deletions, and complex) can be somewhat customized.
- c. Manipulating NHEJ gene expression can control the type and length of indels.
- d. The GEEN binding site may interfere with the binding of key components of classical- or alternative-NHEJ.
- e. Different GEENs may preferentially favor classical- or alternative-NHEJ.

As the interest in gene editing is growing, its impacts on biomedical research, agriculture, gene therapy, and infectious disease control are expected to be vast. Our literature review and analysis identify many different approaches to select or enrich for either insertion or deletion indels. There is clearly experimental support to control the approximate length of insertions and deletions. However, accurately controlling the length of indels may require more research or may not even be possible. More accurate control over indel length could help select for indels in coding regions that are either in-frame or out-of-frame.

There are future improvements that should lead to better control over indel editing. As the NHEJ and gene editing fields focus on the analysis of edits by next-generation sequencing

(NGS) with fluorescent reporter systems, the editing sequence signatures and roles of proteins in repair will become clearer. Reagents for manipulating expression of key NHEJ genes will help researchers select the makeup of indels, perhaps by transient co-transfection with constructs for temporary knockdown by siRNA/shRNA or gene overexpression. As different types of GEENs and architectures emerge, better control over indel features will be achieved.

The interest in indel features and editing signatures may help in the interpretation of indels in genomes that arise from natural phenomena such as retrotransposition, various cancer indels signatures, evolution, and DNA replication. The awareness and out-lined strategies to control indel features herein will likely prove useful for gene editing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Nevada Governor's Office of Economic Development; the National Institutes of Health [R56 AI109156, P20 GM121325]; and the Prabhu Endowed Professorship to M.R.S.

References

- [1]. Gaj T, Gersbach CA, Barbas III C, Trends Biotechnol. 2013, 32, 397.
- [2]. Nemudryi AA, Valetdinova KR, Medvedev SP, Zakian SM, Acta Naturae 2014, 6, 19.
- [3]. Pennisi E, Science 2013, 341, 833. [PubMed: 23970676]
- [4]. Wilkinson R, Wiedenheft B, F1000Prime Rep. 2014, 6, 3. [PubMed: 24592315]
- [5]. Marcaida MJ, Prieto J, Redondo P, Nadra AD, Alibés A, Serrano L, Grizot S, Duchateau P, Pâques F, Blanco FJ, Montoya G, Proc. Natl. Acad. Sci. U.S.A 2008, 105, 16888. [PubMed: 18974222]
- [6]. Stella S, Montoya G, BioEssays 2016, 38, S4. [PubMed: 27417121]
- [7]. Bloom K, Ely A, Mussolino C, Cathomen T, Arbuthnot P, Mol. Ther 2013, 21, 1889. [PubMed: 23883864]
- [8]. Ebina H, Misawa N, Kanemura Y, Koyanagi Y, Sci. Rep 2013, 3, 2510. [PubMed: 23974631]
- [9]. Qu X, Wang P, Ding D, Li L, Wang H, Ma L, Zhou X, Liu S, Lin S, Wang X, Zhang G, Liu S, Liu L, Wang J, Zhang F, Lu D, Zhu H, Nucleic Acids Res. 2013, 41, 7771. [PubMed: 23804764]
- [10]. Strong CL, Guerra HP, Matthey KR, Roy N, Simpson LR, Schiller MR, PLoS One. 2015, 10, e0125652.
- [11]. Arnould S, Delenda C, Grizot S, Desseaux C, Pâques F, Silva GH, Smith J, Protein Eng. Des. Sel 2011, 24, 27. [PubMed: 21047873]
- [12]. Dupuy A, Valton J, Leduc S, Armier J, Galetto R, Gouble A, Lebuhotel C, Sary A, Pâques F, Duchateau P, Sarasin A, Daboussi F, PLoS One 2013, 8, e78678.
- [13]. Mukherjee S, Thrasher AJ, Gene 2013, 525, 174. [PubMed: 23566838]
- [14]. Kim YG, Cha J, Chandrasegaran S, Proc. Natl. Acad. Sci. U.S.A 1996, 93, 1156. [PubMed: 8577732]
- [15]. Guo J, Gaj T, Barbas CF 3rd, J. Mol. Biol 2010, 400, 96. [PubMed: 20447404]
- [16]. Moscou MJ, Bogdanove AJ, Science 2009, 326, 1501. [PubMed: 19933106]
- [17]. Christian M, Cermak T, Doyle E, Schmidt C, Zhang F, Hummel A, Bogdanove A, Voytas D, Genetics 2010, 186, 757. [PubMed: 20660643]
- [18]. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ, Nat. Biotechnol 2011, 29, 143. [PubMed: 21179091]

- [19]. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CVC 3rd, Graveley BR, Terns RM, Terns MP, Mol. Cell 2012, 45, 292. [PubMed: 22227116]
- [20]. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD, Nat. Biotechnol 2013, 31, 822. [PubMed: 23792628]
- [21]. Wei C, Liu J, Yu Z, Zhang B, Gao G, Jiao R, Genet J. Genomics 2013, 40, 281.
- [22]. Iyama T, Wilson DM 3rd, DNA Repair 2013, 12, 620. [PubMed: 23684800]
- [23]. Wyman C, Kanaar R, Annu. Rev. Genet 2006, 40, 363. [PubMed: 16895466]
- [24]. Dueva R, Iliakis G, Trans. Cancer Res 2013, 2, 163.
- [25]. Charbonnel C, Allain E, Gallego ME, White CI, DNA Repair 2011, 10, 611. [PubMed: 21530420]
- [26]. Cortizas EM, Zahn A, Hajjar ME, Patenaude A-M, Di Noia JM, Verdun RE, Immunol J. 2013, 191, 5751.
- [27]. Bétermier M, Bertrand P, Lopez BS, PLoS Genet. 2014, 10, e1004086.
- [28]. Poplawski T, Stoczy ska E, Błasiak J, Postepy Biochem. 2009, 55, 36. [PubMed: 19514464]
- [29]. Bhargava R, Carson CR, Lee G, Stark JM, Proc. Natl. Acad. Sci. U. S. A 2017, 114, 728. [PubMed: 28057860]
- [30]. Jiang G, Plo I, Wang T, Rahman M, Cho JH, Yang E, Lopez BS, Xia F, J. Biol. Chem 2013, 288, 8966. [PubMed: 23344954]
- [31]. Bau D-T, Mau Y-C, Shen C-Y, Cancer Lett. 2006, 240, 1. [PubMed: 16171943]
- [32]. Saha J, Davis AJ, J. Radiat. Res 2016, 57, i18. [PubMed: 27170701]
- [33]. Kijas AW, Lim YC, Bolderson E, Cerosaletti K, Gatei M, Jakob B, Tobias F, Taucher-Scholz G, Gueven N, Oakley G, Concannon P, Wolvetang E, Khanna KK, Wiesmüller L, Lavin MF, Nucleic Acids Res. 2015, 43, 8352. [PubMed: 26240375]
- [34]. Wang H, Rosidi B, Perrault R, Wang M, Zhang L, Windhofer F, Iliakis G, Cancer Res. 2005, 65, 4020. [PubMed: 15899791]
- [35]. Buis J, Stoneham T, Spehalski E, Ferguson DO, Nat. Struct. Mol. Biol 2012, 19, 246. [PubMed: 22231403]
- [36]. Gonzalez KD, Hill KA, Li K, Li W, Scaringe WA, Wang J-C, Gu D, Sommer SS, Hum. Mutat 2007, 28, 69. [PubMed: 16977595]
- [37]. Aza A, Martin MJ, Juarez R, Blanco L, Terrados G, Nucleic Acids Res. 2013, 41, 253. [PubMed: 23143108]
- [38]. Martin MJ, Garcia-Ortiz MV, Gomez-Bedoya A, Esteban V, Guerra S, Blanco L, Nucleic Acids Res. 2013, 41, 9105. [PubMed: 23935073]
- [39]. Chayot R, Montagne B, Ricchetti M, DNA Repair 2012, 11, 22. [PubMed: 22071146]
- [40]. Lieber MR, Annu. Rev. Biochem 2010, 79, 181. [PubMed: 20192759]
- [41]. Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, Páleníková P, Khodak A, Kiselev V, Kosicki M, Bassett AR, Harding H, Galanty Y, Muñoz-Martínez F, Metzakopian E, Jackson SP, Parts L, Nat. Biotechnol 2019, 37, 64.
- [42]. Mann CM, Martínez-Gálvez G, Welker JM, Wiersen WA, Ata H, Almeida MP, Clark KJ, Essner JJ, McGrail M, Ekker SC, Dobbs D, Nucleic Acids Res. 2019, 47, W175. [PubMed: 31127311]
- [43]. Ata H, Ekstrom TL, Martínez-Gálvez G, Mann CM, Dvornikov AV, Schaeffbauer KJ, Ma AC, Dobbs D, Clark KJ, Ekker SC, PLoS Genet. 2018, 14, e1007652.
- [44]. Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, Cassa CA, Liu DR, Gifford DK, Sherwood RI, Nature 2018, 563, 646. [PubMed: 30405244]
- [45]. Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, Cassa CA, Liu DR, Gifford DK, Sherwood RI, Nature 2019, 567, E1. [PubMed: 30765887]
- [46]. Bae S, Kim J-S, Nat. Biomed. Eng 2018, 2, 892. [PubMed: 31015728]
- [47]. Helleday T, Eshtad S, Nik-Zainal S, Nat. Rev. Genet 2014, 15, 585. [PubMed: 24981601]
- [48]. Mullaney JM, Mills RE, Pittard WS, Devine SE, Hum. Mol. Genet 2010, 19, R131. [PubMed: 20858594]
- [49]. Mills RE, Genome Res. 2006, 16, 1182. [PubMed: 16902084]

- [50]. Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu XJ, Ananda G, Howie B, Karczewski KJ, Smith KS, Anaya V, Richardson R, Davis J, The 1000 Genomes Project Consortium, MacArthur DG, Sidow A, Duret L, Gerstein M, Makova KD, Marchini J, McVean G, Lunter G, Genome Res. 2013, 23, 749. [PubMed: 23478400]
- [51]. Viguera E, Canceill D, Ehrlich SD, EMBO J. 2001, 20, 2587. [PubMed: 11350948]
- [52]. Gasior SL, Wakeman TP, Xu B, Deininger PL, J. Mol. Biol 2006, 357, 1383. [PubMed: 16490214]
- [53]. Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y, Carson CT, Macia A, Moran JV, Gage FH, Proc. Natl. Acad. Sci. USA 2011, 108, 20382. [PubMed: 22159035]
- [54]. Suzuki J, Yamaguchi K, Kajikawa M, Ichiyanagi K, Adachi N, Koyama H, Takeda S, Okada N, PLoS Genet. 2009, 5, e1000461.
- [55]. Gasior SL, Roy-Engel AM, Deininger PL, DNA Repair 2008, 7, 983. [PubMed: 18396111]
- [56]. Yamaguchi K, Kajikawa M, Okada N, Mobile Genet. Elem 2015, 5, 92.
- [57]. Vogt J, Bengesser K, Claes KB, Wimmer K, Mautner V-F, van Minkelen R, Legius E, Brems H, Upadhyaya M, Högel J, Lazaro C, Rosenbaum T, Bammert S, Messiaen L, Cooper DN, Kehrer-Sawatzki H, Genome Biol. 2014, 15, R80. [PubMed: 24958239]
- [58]. Vu GTH, Cao HX, Watanabe K, Hensel G, Blattner FR, Kumlehn J, Schubert I, Plant Cell 2014, 26, 2156. [PubMed: 24876253]
- [59]. Uehara Y, Ikehata H, Komura J, Ito A, Ogata M, Itoh T, Hirayama R, Furusawa Y, Ando K, Paunesku T, Woloschak GE, Komatsu K, Matsuura S, Ikura T, Kamiya K, Ono T, Radiat. Res 2008, 170, 216. [PubMed: 18666816]
- [60]. Boubakour-Azzouz I, Bertrand P, Claes A, Lopez BS, Rougeon F, Nucleic Acids Res. 2012, 40, 8381. [PubMed: 22740656]
- [61]. Ma Y, Lu H, Tippin B, Goodman MF, Shimazaki N, Koiwai O, Hsieh C-L, Schwarz K, Lieber MR, Mol. Cell 2004, 16, 701. [PubMed: 15574326]
- [62]. Capp J-P, Boudsocq F, Bertrand P, Laroche-Clary A, Pourquier P, Lopez BS, Cazaux C, Hoffmann J-S, Canitrot Y, Nucleic Acids Res. 2006, 34, 2998. [PubMed: 16738138]
- [63]. Dominguez O, EMBO J. 2000, 19, 1731. [PubMed: 10747040]
- [64]. Ramadan K, Maga G, Shevelev IV, Villani G, Blanco L, Hübscher U, J. Mol. Biol. 2003, 328, 63. [PubMed: 12683997]
- [65]. Kusumoto R, Dawut L, Marchetti C, Wan Lee J, Vindigni A, Ramsden D, Bohr VA, Biochemistry 2008, 47, 7548. [PubMed: 18558713]
- [66]. Lee-Theilen M, Matthews AJ, Kelly D, Zheng S, Chaudhuri J, Nat. Struct. Mol. Biol 2011, 18, 75. [PubMed: 21131982]
- [67]. Oh S, Harvey A, Zimbric J, Wang Y, Nguyen T, Jackson PJ, Hendrickson EA, DNA Repair 2014, 21, 97. [PubMed: 24837021]
- [68]. Mashimo T, Kaneko T, Sakuma T, Kobayashi J, Kunihiro Y, Voigt B, Yamamoto T, Serikawa T, Sci. Rep 2013, 3, 10.1038/srep01253.
- [69]. Rass E, Grabarz A, Plo I, Gautier J, Bertrand P, Lopez BS, Nat. Struct. Mol. Biol 2009, 16, 819. [PubMed: 19633668]
- [70]. Fattah F, Lee EH, Weisensel N, Wang Y, Lichter N, Hendrickson EA, PLoS Genet. 2010, 6, e1000855.
- [71]. Xie A, Kwok A, Scully R, Nat. Struct. Mol. Biol 2009, 16, 814. [PubMed: 19633669]
- [72]. Jantz D, Berg JM, Biophys. J 2010, 98, 852. [PubMed: 20197039]
- [73]. Beumer KJ, Trautman JK, Christian M, Dahlem TJ, Lake CM, Hawley RS, Grunwald DJ, Voytas DF, Carroll D, G3: Genes, Genomes, Genet 2013, 3, 1717.
- [74]. Grabher C, Wittbrodt J, Genome Biol. 2007, 8, S10. [PubMed: 18047687]
- [75]. Wende W, Grindl W, Christ F, Pingoud A, Pingoud V, Nucleic Acids Res. 1996, 24, 4123. [PubMed: 8932361]
- [76]. Wolfs JM, DaSilva M, Meister SE, Wang X, Schild-Poulter C, Edgell DR, Nucleic Acids Res. 2014, 42, 8816. [PubMed: 25013171]

- [77]. Eastberg JH, McConnell Smith A, Zhao L, Ashworth J, Shen BW, Stoddard BL, Nucleic Acids Res. 2007, 35, 7209. [PubMed: 17947319]
- [78]. Mueller JE, Smith D, Belfort M, Genes Dev. 1996, 10, 2158. [PubMed: 8804310]
- [79]. Westra ER, Semenova E, Datsenko KA, Jackson RN, Wiedenheft B, Severinov K, Brouns SJJ, PLoS Genet. 2013, 9, e1003742.
- [80]. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA, Nature 2014, 507, 62. [PubMed: 24476820]
- [81]. Gasiunas G, Barrangou R, Horvath P, Siksnys V, Proc. Natl. Acad. Sci. U. S. A 2012, 109, E2579.
- [82]. Bolukbasi MF, Liu P, Luk K, Kwok SF, Gupta A, Amrani N, Sontheimer EJ, Zhu LJ, Wolfe SA, Nat. Commun 2018, 9, 4856. [PubMed: 30451839]
- [83]. Wolfs JM, Hamilton TA, Lant JT, Laforet M, Zhang J, Salemi LM, Gloor GB, Schild-Poulter C, Edgell DR, Proc. Natl. Acad. Sci. U.S.A 2016, 113, 14988.
- [84]. Ye K, Wang J, Jayasinghe R, Lameijer E-W, McMichael JF, Ning J, McLellan MD, Xie M, Cao S, Yellapantula V, Huang K, Scott A, Foltz S, Niu B, Johnson KJ, Moed M, Slagboom PE, Chen F, Wendl MC, Ding L, Nat. Med 2016, 22, 97. [PubMed: 26657142]
- [85]. Kim Y, Kweon J, Kim J-S, Nat. Methods 2013, 10, 185. [PubMed: 23396284]
- [86]. Chuzhanova NA, Anassis EJ, Ball EV, Krawczak M, Cooper DN, Hum. Mutat 2003, 21, 28. [PubMed: 12497629]
- [87]. Richardson CD, Ray GJ, Bray NL, Corn JE, Nat. Commun 2016, 7, 12463.
- [88]. Christian ML, Demorest ZL, Starker CG, Osborn MJ, Nyquist MD, Zhang Y, Carlson DF, Bradley P, Bogdanove AJ, Voytas DF, PLoS One 2012, 7, e45383.
- [89]. Kosicki M, Rajan SS, Lorenzetti FC, Wandall HH, Narimatsu Y, Metzakopian E, Bennett EP, Prog. Mol. Biol. Transl. Sci 2017, 152, 49. [PubMed: 29150004]
- [90]. Walker E, Hernandez AV, Kattan MW, Cleveland Clin. J. Med 2008, 75, 431.
- [91]. Hjelm LN, Chin ELH, Hegde MR, Coffee BW, Bean LJH, J. Mol. Diagn 2010, 12, 607. [PubMed: 20639189]
- [92]. Finnie NJ, Gottlieb TM, Blunt T, Jeggo PA, Jackson SP, Proc. Natl. Acad. Sci. U. S. A 1995, 92, 320. [PubMed: 7816841]
- [93]. Meek K, Kienker L, Dallas C, Wang W, Dark MJ, Venta PJ, Huie ML, Hirschhorn R, Bell T, Immunol J. 2001, 167, 2142.
- [94]. Huertas P, Nat. Struct. Mol. Biol 2010, 17, 11. [PubMed: 20051983]
- [95]. Piganeau M, Ghezraoui H, De Cian A, Guittat L, Tomishima M, Perrouault L, René O, Katibah GE, Zhang L, Holmes MC, Doyon Y, Concordet J-P, Giovannangeli C, Jasin M, Brunet E, Genome Res. 2013, 23, 1182. [PubMed: 23568838]
- [96]. Cotner-Gohara E, Kim I-K, Hammel M, Tainer JA, Tomkinson AE, Ellenberger T, Biochemistry 2010, 49, 6165. [PubMed: 20518483]
- [97]. Gouge J, Rosario S, Romain F, Beguin P, Delarue M, J. Mol. Biol 2013, 425, 4334. [PubMed: 23856622]
- [98]. Orans J, McSweeney EA, Iyer RR, Hast MA, Hellinga HW, Modrich P, Beese LS, Cell 2011, 145, 212. [PubMed: 21496642]
- [99]. Ali AAE, Timinszky G, Arribas-Bosacoma R, Kozlowski M, Hassa PO, Hassler M, Ladurner AG, Pearl LH, Oliver AW, Nat. Struct. Mol. Biol 2012, 19, 685. [PubMed: 22683995]
- [100]. Kitano K, Kim S-Y, Hakoshima T, Structure 2010, 18, 177. [PubMed: 20159463]
- [101]. Walker JR, Corpina RA, Goldberg J, Nature 2001, 412, 607. [PubMed: 11493912]

Figure 1. Types of gene editing endonuclease technologies. A schematic representation of different GEEN technologies with DNA binding sites and spacer regions indicated. The sequence in capital font indicates DNA binding sites and arrows indicate approximate cleavage sites. Sequences are colored alternating grey in increments of 5 bp.

Figure 2.

Indel deletion/insertion ratios. Bar plot showing the ratio of deletions/insertions for environmental damage, different gene technologies and editing in cells containing null alleles for different genes as indicated. The black bar indicates the control. The bar plot is calculated as the ratio of the sum of the deletion edits over the insertion edits for each condition represented on the X-axis.

Figure 3.

Editing signatures of null alleles for NHEJ genes. Bar charts of editing profiles for cell harboring null alleles of genes in the C-NHEJ pathway for insertions (A) and deletions (B) and in the A-NHEJ pathway for insertions (C) and deletions (D). Keys indicate which genes are deleted and numbers in parentheses indicate the number of edit events. “-” denotes knockout, “-” denotes knockdown, and “+” denotes overexpression.

Figure 4. Summary of approaches to control indel makeup. A) Types of Indels generated (deletions, insertions and complex) after DSB repair. B) A schematic diagram summarizing indel lengths (insertions and deletions) constituted by knockout (Δ), knockdown (–) and over expression (+) of color-coded C-NHEJ genes (blue), A-NHEJ genes (grey), and GEENs (purple).

Figure 5.

NHEJ protein binding sites on double-stranded DNA. The relative spacing of key recruitment factors for C-NHEJ (A), A-NHEJ (B). The lengths of the DNA binding sites and spaces are indicated. Blue lines indicate spacing of the approximate binding sites for TALENs/ZFNs and orange lines indicate the approximate binding sites for Meganuclease and CRISPR/Cas9. Binding sites are from structures of KU70 (1JEY), and PARP1 (4AV1); the length of the DNA binding site and distance from the end are indicated. PARP1 binds to the end. Not shown are WRN (3AAF) binding 2 bp at the end of the DNA and Exo1 binding

to the last 9 bp (PDB Accession number: 3QEB), TdT binding the last 5 bp of single-stranded DNA (4I29), and LIGIII binding the last 6–7 bp (3L2P).[96–101] C) Bar plot showing the percentage of edits that include all or part of the DNA binding site of the specific GEEN (grey bars) and the percentage of edits with complex indels (black bars).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 6.

Summary of environmental induced DSBs and breaks from GEEN editing of genome. Editing profiles show the percentage and lengths of insertions (grey bars) and deletions (black bars) based on experiments for A) environmental damage ($n = 180$ repair events); and editing by B) Meganucleases ($n = 366$ repair events); C) ZFNs ($n = 145$ repair events); D) TALENs ($n = 676$ repair events); and E) CRISPR/Cas9 ($n = 999$ repair events).