A genome editing primer for the hematologist

Megan D. Hoban and Daniel E. Bauer

Introduction

The immense potential of site-specific genomic modification has been appreciated since the advent of molecular genetics. Prior to the availability of engineerable targeted nucleases, homologous recombination could be instigated in mammalian cells by the presence of extrachromosomal donor templates, including small oligonucleotides. The uses of triple helix forming oligonucleotides as triggers and viral genomes such as adeno-associated virus (AAV) as donors for homologous recombination in the absence of additional sequences.1-4 The uses of triple helix forming oligonucleotides as triggers and viral genomes such as adeno-associated virus (AAV) as donors for homologous recombination in the absence of additional sequences.1-4

Gene editing enables the site-specific modification of the genome. These technologies have rapidly advanced such that they have entered common use in experimental hematology to investigate genetic function. In addition, genome editing is becoming increasingly plausible as a treatment modality to rectify genetic blood disorders and improve cellular therapies. Genome modification typically ensues from site-specific double-strand breaks and may result in a myriad of outcomes. Even single-strand nicks and targeted biochemical modifications that do not permanently alter the DNA sequence (epigenome editing) may be powerful instruments. In this review, we examine the various technologies, describe their advantages and shortcomings for engendering useful genetic alterations, and consider future prospects for genome editing to impact hematology. (Blood. 2016;127(21):2525-2535)

Genome editing tools

Meganucleases

Meganucleases (also called homing endonucleases) were the first targeted nuclease described.6 These 20- to 37-kDa (~0.6-1 kb) sequence-specific nucleases are active as monomers and named based on their ability to recognize relatively long (14-40 bp) target sequences.7 Meganucleases are distinguished from restriction endonucleases by their much larger target site size, whose occurrence may be as rare as a single instance per genome.7 Applications of meganucleases are limited by the low frequency of target site presence at most genes, thus essentially preventing the use of naturally occurring forms for functional genomics or therapeutic development.10 Meganucleases may be engineered by both rational structure-guided design and high-throughput screening to recognize diverse targets, including those within human genes, although this remains a nontrivial task.18 In addition, the tendency of meganucleases to tolerate some target sequence degeneracy may limit their specificity.21

Zinc-finger nucleases

Zinc-finger nucleases (ZFNs) are chimeric nucleases that consist of individual zinc-finger protein (ZFP) motifs, each of which is capable of specifically recognizing 3 to 4 bases of DNA.22 Linking multiple ZFPs together allows for the targeting of extended stretches of DNA (typically 9-18 bp). The coupling of a FokI nuclease (which cleaves
without sequence specificity but requires dimerization) to each of a pair of monomers allows the composite dimeric enzyme to introduce a site-specific DSB. Each ZFN target site is composed of 2 ZFP binding sites on either side of a spacer region of 5 to 7 bp within which the dimerized FokI cleaves. Although the need for dimerization allows for increased target site specificity (due to the increased binding site size), wild-type FokI can still homodimerize and lead to undesired breaks. FokI domains containing mutations (ELD/KKR) in the dimerization interface result in increased specificity via obligate heterodimerization. Each ZFN monomer is encoded by a roughly 1.2 kb (~45 kDa) gene, a pair of which need to be delivered to target cells to execute genome editing. Like meganucleases, the requirement for de novo design and testing of ZFN target sequence recognition is a barrier to their widespread use.

Transcription activator–like effector nucleases

Similar to ZFNs, transcription activator–like effector nucleases (TALENs) are chimeric nucleases composed of an engineerable DNA-binding domain and a FokI nuclease. The DNA binding domain consists of an array of individual TALE repeat modules composed of 34 amino acid modules that determine DNA specificity based merely on the sequence of the 12th and 13th residues (so-called repeat variable diresidues [RVDs]). Like ZFNs, TALENs are typically designed to act as obligate heterodimers to increase specificity. Thus, active TALENs consist of a pair of monomers (2.5-3.0 kb or 90-110 kDa) separated by a 12- to 20-bp spacer region that introduce a DSB on FokI dimerization. TALEN design and assembly is more straightforward than that of ZFNs as each RVD targets a single DNA base and multiple RVDs can be stitched together in essentially any order. Online resources are available to assist with design and cloning from preformed modules. Hybrid nucleases consisting of the easily engineerable TALE binding domain and the site-specific meganuclease head (so-called megaTALs) may offer additional specificity, affinity, and potentially simplified cellular delivery because they are active as ~75-kDa (~2-kb) monomers.

Clustered regularly interspaced short palindromic repeat/Cas

Clustered regularly interspaced short palindromic repeat (CRISPR) sequences are a defining component of a prokaryotic adaptive immune system. From these loci, bacteria express genomically encoded RNAs to guide nuclease cleavage to matching sequences of invading phage and plasmid DNA. The widely studied type II system from Streptococcus pyogenes consists of 3 components: (1) the nuclease Cas9, containing both RuvC and HNH nuclease domains, which produces blunt DSBs; (2) a DNA-binding CRISPR RNA (crRNA), which includes a 20-nt guide RNA (gRNA) sequence with precise complementarity to its DNA target; and (3) an auxiliary trans-activating crRNA (tracrRNA) bridging the crRNA to Cas9. Recognition of a target site by SpCas9 depends on the presence of a protospacer adjacent motif (PAM) sequence NGG immediately downstream of the gRNA target sequence. Cleavage occurs 3 bp upstream of the PAM.

The most common implementation of CRISPR genome editing in eukaryotic cells relies on a 2-component system consisting of SpCas9 (~160 kDa, ~4.2 kb) and a single chimeric guide RNA (sgRNA) of ~110 nt that subsumes the recognition and structural functions of the crRNA and tracrRNA.

The discrete rules underlying SpCas9/sgRNA genome editing belie the extensive diversity of CRISPR/Cas systems. Several naturally occurring Cas9 orthologs have been adapted for genome editing of human cells, and many more are likely to be discovered. For example, the smaller size of the Staphylococcus aureus Cas9 (~120 kDa, ~3.3 kb) compared with SpCas9 may provide advantages in terms of cellular delivery, but its restriction by a different PAM sequence (NNGRRT) may limit SaCas9 to a somewhat narrower set of genomic targets. Engineered variants of SaCas9 and SpCas9 with novel PAM recognition sequences broaden the repertoire of possible genomic targets. Cpf1, a member of a distantly related CRISPR nuclease family, differs from SpCas9 in that it uses a single gRNA without tracrRNA, has a different PAM sequence (TTN), appears to act as a dimer rather than monomer, and introduces a staggered rather than blunt DSB. This multitude of CRISPR/Cas variants provides scientists with a rapidly expanding toolkit with extensive versatility.

Molecular outcomes of genome editing

Following introduction of a DSB in the genome, cellular repair prevents disastrous chromosomal catastrophes. The varied repair responses may be simplified into 2 major pathways, each of which relies on a large number of host factors: nonhomologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ involves reuniting the 2 broken ends of the chromosome in a process that often leads to small insertions or deletions (so-called indels) at the cleavage site. These small “scars” left behind by indel mutations may result in functional disruption of essential target sequences, such as producing missense or frameshift mutations or interruption of splice sites or transcription factor binding sites (Figure 2A). The simultaneous introduction of 2 DSBs some distance away from each other on the same chromosome may result in large interstitial deletions and inversions or in translocations if on different chromosomes (Figure 2B).

In contrast, HDR depends on a donor template for repair. In “natural” HDR, following genotoxic injury, homologous sister chromatids serve as template for precise repair in replicating cells. In the case of genome editing, an extrachromosomal donor sequence may be used to integrate sequences of choice adjacent to induced DSBs (or even to single-strand breaks, as may be produced by nickase mutants of Cas9). Because expression of HDR pathway components is limited to the S/G2 phase of the cell cycle, only dividing cells are competent for this type of repair.

Genome editing as a research tool

A fundamental goal of human genetics is to relate genotype to phenotype. Genome editing provides a powerful means to rapidly generate novel alleles to determine the function of coding and noncoding sequences. Given its robustness and ease of use, CRISPR/Cas9 has quickly become an indispensable tool for many experimental hematologists to evaluate gene function. The ability to create null alleles on demand appears at least as powerful a paradigm shift as preceding RNA interference technology for functional genomics. Various manifestations of genome editing as a research methodology include the production of knockout and knock-in alleles in hematopoietic cell lines and in hematopoietic stem cells transplanted to mice, to model leukemia-associated translocations, and to produce various animal models of genetic modification in both the zygote and restricted to somatic lineages.

The control of Cas9 specificity merely by the nature of the 20-nt gRNA sequence makes the generation and interrogation of large-scale sgRNA libraries straightforward by massively parallel oligonucleotide
synthesis and sequencing, respectively. With increased specificity and efficiency over RNAi screens, CRISPR-based forward genetic screens may be performed by introduction to cells of genome-wide libraries of sgRNAs as mutagens, cells with a given phenotype retrieved, and then the mutagens of interest easily identified.\textsuperscript{52,53} These approaches have been successfully applied to identify unexpected dependencies for various infectious diseases and cancers.\textsuperscript{54-56}

Systematic CRISPR approaches not only provide an opportunity to identify genes of interest, but also delineate key coding and non-coding sequences. Dense sgRNA libraries targeted throughout coding sequences can identify functional domains within a gene, such as druggable dependencies in hematologic malignancy.\textsuperscript{57} Methodical perturbation of regulatory sequences, such as the fetal hemoglobin-associated erythroid-specific enhancer of \textit{BCL11A}\textsuperscript{58} by CRISPR/Cas9 and ZFN/TALEN mutagenesis, have identified critical regulatory sequences that themselves may serve as therapeutic targets.\textsuperscript{59,60}

In contrast to conventional genome editing, which implies permanent modification of genomic target sequences, genome editing tools may be repurposed to engender potent biological outcomes without mutagenesis. For example, catalytically inactive mutants of Cas9 (dCas9) may be guided to proximal promoters or coding sequences to block transcription.\textsuperscript{61} Linking the site-specific binding domains of targeted nucleases to gene regulatory domains (along with disabled nuclease function) may allow for potent modulation of gene expression to characterize mechanisms of transcriptional regulation as well as redirect cellular phenotype.\textsuperscript{62} For example, fusion of dCas9 to
a repressor such as the Kruppel-associated box promotes repressive histone modifications such as histone H3K9 trimethylation and targeted gene silencing. Pairing TALEs or dCas9 with chromatin regulators such as LSD1 histone demethylase can result in targeted epigenetic regulation of noncoding elements and help define relationships between genes and distal regulatory elements.

In addition to gene repression, gene activation can also be achieved using modified targeted nuclease platforms. Coupling dCas9 to individual positive transcriptional regulators such as the VP64 transactivation domain or the core region of the histone acetyltransferase p300 results in robust transcription of target genes, and higher-order clusters of activation domains may serve as even more potent activators. Targeting of such dCas9-based transcriptional activators or repressors to a large set of gene promoters with sgRNA libraries permits genome-wide screens.

Therapeutic considerations

Repair mode

The ultimate cure for a genetic disorder would be to revert the disease mutation back to normal sequence. Conditions for which a single or predominant mutation underlies the disease would seem to be most amenable to this approach (because a single targeting strategy could be broadly applicable). In addition, a point mutation might be simpler to correct than a larger scale rearrangement. For example, sickle cell disease, with its high incidence and characteristic causative β-globin adenine-to-thymine transversion, would seem an auspicious target for gene correction. In fact, HDR-based genome editing correction of the sickle cell disease point mutation has been achieved in multiple cell types, including induced pluripotent stem cells, although the rates of editing that could be achieved in long-term hematopoietic stem cells (HSCs) have thus far remained below the level of therapeutic relevance.

For other conditions with numerous mutations of an affected gene, like the varied mutations of IL2RG underlying X-linked severe combined immunodeficiency (X-SCID), a more universal strategy might be attractive, such as the targeted integration of an entire gene cassette, although the rates of editing that could be achieved in long-term hematopoietic stem cells (HSCs) have thus far remained below the level of therapeutic relevance.

Figure 2. Molecular outcomes of genome editing.

Multiple possible repair outcomes of (A) 1 or (B) 2 DSBs in the genome. Schematics of indels, gene insertions, and gene correction are shown. Solid gold lines, inserted bases; dashed red lines, deleted bases; solid dark blue boxes, gene cassettes. Outcomes depicted are not exhaustive, and other outcomes, such as insertion of a gene cassette via NHEJ, are possible.
conditions in which endogenous gene regulation is imperative. For example, CD40 ligand (CD40L), whose deficiency results in combined immunodeficiency, is typically only transiently expressed on activated T cells. Constitutive expression of CD40L in T cells of CD40L-deficient mice results in lymphoproliferative disorder, suggesting the potential risks of nonendogenous gene regulation.\textsuperscript{77} Although targeted integration often insinuates HDR repair, NHEJ pathways may also be harnessed to achieve targeted integration, which may be particularly important in nondividing cells.\textsuperscript{78-80} Even if a mutant gene were effectively replaced or replaced, an additional challenge may be to avoid immune responses against the novel gene product.\textsuperscript{91}

Although gene correction might seem the most intuitive approach to therapeutic genome editing, in fact the first clinical trial using targeted nucleases in human patients has relied on NHEJ-based genetic disruption rather than gene repair. One advantage of this strategy is that NHEJ tends to be a more active repair pathway compared with HDR, particularly in quiescent cells.\textsuperscript{82} Another benefit of gene disruption may be that modulation of a disease-modifying pathway could produce a more universal remedy for the disease compared with individual mutation-specific corrections. The first-in-human genome editing trial using ZFNs in autologous T cells to target the HIV5 coreceptor CCR5,\textsuperscript{83} inspired by individuals naturally resistant to HIV infection due to CCR5 deficiency, as well as the remarkable case of an HIV-positive recipient of an allogeneic CCR5-deficient hematopoietic stem cell transplant (HSCT) whose HIV became undetectable without antiretroviral therapy.\textsuperscript{56,84}

The indels left behind following NHEJ repair may be useful not only for disrupting coding sequences but also noncoding regulatory sequences such as the erythroid-specific enhancer of BCL11A for the β-hemoglobinopathies\textsuperscript{59,60} or DMD splice sites to promote exon skipping in muscular dystrophy.\textsuperscript{85} Paired DSBs may also result in desirable outcomes, such as interstitial deletions of mutant exons in DMD\textsuperscript{85-87} or reversion of large chromosomal inversions underlying hemophilia A.\textsuperscript{44}

Even epigenetic editing approaches could be considered for therapeutic utility, such as forced chromatin looping for the β-hemoglobinopathies to redirect interactions with the powerful globin enhancer locus control region away from mutant β-globin toward compensatory γ-globin.\textsuperscript{88,89} To the degree that persistent expression of the artificial DNA binding factor would be required to maintain therapeutic gene expression, the kiss-and-run promise of therapeutic genome editing would be lost, and the challenges of long-term stable and safe expression would be similar to those facing conventional gene therapy.\textsuperscript{90}

### Target cell

The ideal somatic target cell for therapeutic gene editing depends on the specific clinical situation. In general, the goal is for a durable, potentially even lifelong (ie, curative), benefit from gene editing. The ability to access and manipulate the appropriate target cell may determine the feasibility of the endeavor. Limiting edits to disease-associated cells may minimize risk of adverse clinical effects of genetic perturbation of cell types unconnected to the disease, although the nature of this risk depends on the particular perturbation.

For many hematologic conditions, the most relevant cell type to edit would be the HSC, the rare self-renewing cells atop the hematopoietic hierarchy. Certainly for genetic disorders of the HSC itself (such as inherited bone marrow failure disorders like Fanconi anemia or dyskeratosis congenita), modification of this cell would be required for a salutary effect. For monoclonic disorders of downstream lineages, such as of erythrocytes (eg, hemoglobinopathies\textsuperscript{91}) or granulocytes (eg, chronic granulomatous disease\textsuperscript{92,93}), modification of the HSC appears to be required to achieve a renewing source of corrected cells. Engineering HSCs may even have the advantage to confer exceptional properties (ie, more ameliorating than mere healthy cells) to downstream progeny. For example, supraphysiologic expression of asrylsulfatase A in corrected microglia derived from autologous HSCs in metachromatic leukodystrophy appears to confer benefits beyond that of allogeneic HSCT due to cell nonautonomous cross-correction.\textsuperscript{94} For all HSC-based therapies, in addition to any risks intrinsic to the gene editing itself, risks of isolating the cells from the appropriate source and preparative conditioning therapy must be considered.

One caveat to engineering hematopoiesis is that most knowledge of HSC function is based on cellular capacity to reconstitute ablative animals and patients. Emerging studies of steady-state hematopoiesis suggest that a large fraction of hematopoietic output derives from long-lived lineage-restricted progenitor cells,\textsuperscript{95-97} raising the question of whether modifying appropriate lineage-restricted progenitors could have durable therapeutic utility under specific conditions.

For inherited disorders of lymphocytes (eg, immunodeficiencies), modification of long-lived T cells might sometimes be adequate, although in some conditions, T cells may be absent. The results of gene therapy for X-SCID suggest that modification of a barely measurable fraction of HSCs can result in robust (although sometimes oligoclonal) T-cell reconstitution, reflecting the intense selective advantage for rescued T cells in certain immunodeficiencies.\textsuperscript{98} Another consideration is the number of rescued cells required to provide a therapeutic advantage. For example, in CD40L deficiency, allogeneic HSCT with mixed chimerism may be curative.\textsuperscript{99} Therefore, in this type of condition, correction of merely a subset of HSCs or even T cells might be expected to be of therapeutic value.

Isolated T-cell modification may be desirable in various adoptive immunotherapy applications. A technical advantage may be that T cells can be easier to expand and manipulate ex vivo compared with HSCs\textsuperscript{100} (Figure 3). Also T-cell editing could have a theoretical safety advantage with respect to myeloproliferative risk. Allogeneic T-cell therapies could be useful in a wide variety of immune, infectious, and malignant conditions. Autologous T cells with enforced expression of chimeric antigen receptors (CARs),\textsuperscript{101-103} recognizing tumor-associated antigens have shown remarkable clinical responses; however, the current approaches require expensive, labor-intensive, time-sensitive autologous cell processing. Gene disruption of the endogenous T-cell receptor (TCR) might allow for production of allogeneic CAR T cells while avoiding risk of suboptimal or graft-versus-host responses.\textsuperscript{104,107} Recently, a report of a single remarkable case with molecular remission following treatment with TALEN TCR-edited allogeneic CAR T cells (with subsequent allogeneic HSCT) for multiply relapsed B-cell acute lymphoblastic leukemia has generated great excitement.\textsuperscript{108}

Other long-lived hematopoietic cells in addition to HSCs might be considered as targets for therapeutic gene editing. For example, congenital pulmonary alveolar proteinosis results from absence of pulmonary alveolar macrophages due to granulocyte-macrophage colony-stimulating factor deficiency. In mouse models, transplantation of gene corrected or healthy pulmonary alveolar macrophages can result in disease reversal for ≥1 year,\textsuperscript{109} suggesting that tissue macrophages can be genetically manipulated, long-lived, and biologically potent under particular clinical circumstances.

A variety of hematologic conditions might benefit from gene editing of nonhematopoietic lineages, chiefly the inherited bleeding disorders. Hepatocytes are both the frequent target for various genetic therapies\textsuperscript{110} and the natural source of synthesis of various coagulation factors that may be congenitally deficient such as factor IX in hemophilia B. However, expression of factor IX from cells besides hepatocytes (eg, myocytes, megakaryocytes\textsuperscript{111,112}) and expression of factor VIII from
hepatocytes (rather than endogenous endothelial cells) may be restorative in the hemophilias. These results suggest that recapitulating the endogenous cell source for secreted factors may be less important than other considerations such as ease and safety of delivery and achievable level of transgene expression.

Finally, genome editing of pluripotent stem cells has a number of advantages, including the ability to select rare desired mutations and to extensively characterize clonal on-target and off-target effects, including by whole genome sequencing. The major challenge is to develop efficient protocols to produce hematopoietic stem cells or other clinically relevant cellular outputs. Irrespective of the target cell, a successful clinical gene editing approach must consider the fraction of the tissue that must be corrected to achieve therapeutic benefit. This may depend on the degree of selective advantage of effector cell relative to target cell (eg, survival advantage to be expected for erythrocytes and erythroid precursors deriving from corrected HSCs in hemoglobinopathy or for B-lymphocytes derived from corrected HSCs in X-linked agammaglobulinemia).

**Delivery**

Robust cellular delivery may be the limiting factor for therapeutic genome editing. Most strategies involve just a transient burst of nuclease expression with the rationale that after desired edits have been produced, nucleases serve no productive role. Meganucleases and megaTALs need only the expression of 1 protein, ZFNs and TALENs require the expression of 2 monomers, and the CRISPR/Cas9 system depends on simultaneous expression of both a protein (Cas9) and RNA (sgRNA) to yield an active ribonucleoprotein complex in target cells.

For genome editing of hematopoietic cells, most attention has focused on ex vivo delivery. TALENs have proven difficult to deliver via AAV or lentiviral vectors due to their large size and highly repetitive sequences. Delivery of SpCas9 via AAV is particularly challenging given that the gene itself nearly outstrips the viral genome cargo limit, leaving very little room for regulatory elements, sgRNA cassettes, or homology donor sequences. Smaller orthologs such as SaCas9 appear better suited for AAV delivery. Concerns of low-level AAV integration could become particularly relevant in the setting of clinical delivery of active nucleases. Although widely tropic integrase-defective lentiviral vectors might appear a logical approach, these are hindered by relatively low gene expression levels compared with integrating counterparts.

Transient delivery of targeted nucleases to primary hematopoietic cells has been most successfully achieved by electroporation. Interestingly, electroporation of mRNA has reduced risk of integration and appears to result in reduced cellular toxicity compared with DNA. Delivery of in vitro transcribed ZFN mRNA to HSCs and T cells via electroporation may be compatible with clinical-scale genome editing. CRISPR/Cas9 may be delivered to both HSCs and T cells via electroporation of ribonucleoprotein complexes of Cas9 protein and sgRNA.

Alternative delivery platforms such as osmocytosis, cell-penetrating peptides, cationic lipids, and microfluidic devices have been proposed, but their efficacy and lack of toxicity for primary hematopoietic cells remain to be determined.

In contrast to the focus on ex vivo delivery for hematopoietic cells, in vivo delivery has been investigated particularly for therapeutic genome editing of nonhematopoietic tissues. For example, intravenous delivery by AAV of ZFNs and a donor cassette targeting the factor IX locus allows for the targeted integration and therapeutic production of factor IX in hemophilia B mice. An alternate strategy is to use AAV delivery of ZFNs and a promoterless donor cassette to target expression to the first intron of the albumin locus. Recent proof-of-principle for this approach has been demonstrated to ameliorate both factor VIII and factor IX deficiencies in relevant mouse models.
In addition, 3 groups recently showed that AAV delivery of Cas9 and a pair of sgRNAs could delete a mutant exon, resulting in therapeutic exon skipping in mouse models of Duchenne muscular dystrophy.\textsuperscript{85-87} Collectively, these studies demonstrate proof-of-principle of systematic delivery of nucleases to target genome editing in liver and muscle, each of which might serve as appropriate sites of synthesis in coagulation factor deficiency.

In addition to viral vectors, nonviral delivery, particularly to the liver, may be considered for in vivo genome editing. Already hydrodynamic injection of CRISPR/Cas9 plasmids has been shown to correct the metabolic liver disease hereditary tyrosinemia in mice.\textsuperscript{134} Other forms of nonviral in vivo delivery, such as ultrasound with microbubbles, cationic lipid complexes, and DNA nanoparticles, have been investigated.\textsuperscript{135-137} Given the pace of advances, even in vivo delivery of genome editing reagents targeted directly to hematopoietic cells including HSCs seems conceivable. In contrast to their typical use via ex vivo transduction, lentiviral and nonlentiviral vectors may directly transduce HSCs following in vivo intravenous or intrasosseous delivery.\textsuperscript{138,139} One issue that in vivo delivery might face is development of immune responses against the targeted nucleases themselves (ie, bacterially derived nucleases in the case of CRISPR), which could restrict the potential for successive therapies.

Editing efficiency and specificity

The ideal genome edit would occur with high efficiency at the desired ("on-target") site and with few if any consequences at other ("off-target") genetic loci.\textsuperscript{140} Both efficiency and specificity of editing may depend on the nature of the nuclease platform itself, features of the genomic target site, and the mechanisms by which the nucleases are delivered to the target cells.

Various claims have been made about the relative superiority of numerous platforms.\textsuperscript{141} Overall, it is unclear whether differences observed in individual studies are true class effects or related to the individual loci, target sequences, cell types, and delivery modalities. Particularly given the numerous permutations of CRISPR/Cas systems, it is difficult to generalize. For an individual nuclease, its level of expression appears to be a critical determinant of efficiency.\textsuperscript{142}

Cell type–specific chromatin context may play an important role in determining accessibility and subsequent modification of a given nuclease at a target site.\textsuperscript{143,144} Typically alleles repaired by NHEJ exceed those repaired by HDR following introduction of DSBs, particularly in nondividing cells. A variety of strategies have been used to promote HDR compared with NHEJ outcomes, including pharmacologic or genetic inhibition of NHEJ, augmentation of HDR, and synchronization of cell cycle.\textsuperscript{145,146} Some have suggested that the unique 3′ overhang generated by a meganuclease head (including by megaTALs) may enhance the rate of HDR relative to FokI or Cas9 cleavage.\textsuperscript{100}

Gene therapy trials with integrating retroviral vectors targeting HSCs have demonstrated both clinical efficacy and severe adverse events due to vector integration-mediated oncogenesis, although newer-generation lentiviral vectors appear to substantially reduce risk of insertional mutagenesis.\textsuperscript{147-149} In contrast to conventional gene therapy, which counts as success thousands of unique semirandom vector integration events with many distributed near active genes, genome editing raises the theoretical potential for seamless repair, with only the desired genetic modification. Numerous strategies have been described to minimize risks of off-target mutagenesis in response to targeted nuclease exposure.

The first step is to design reagents with the lowest possible risk for off-target cleavages. Target sequences with exact or close matches in the genome in addition to the on-target site carry the greatest risk for off-target cleavage because targeted nuclease DNA recognition may be somewhat promiscuous. For CRISPR gRNAs, the so-called seed sequences proximal to the PAM are particularly intolerant of mismatches.

Second, once any close matching sequences (possible off-target sites) are identified, their cleavage may be closely monitored experimentally.\textsuperscript{150,152} However, the ability to predict off-target sites may be imperfect, indicating the importance of systematic approaches for unbiased assessment of off-target cleavage.\textsuperscript{153,154} Whole-genome sequencing may be considered the ultimate unbiased technique, but the sensitivity is relatively low and thus unable to exclude rare off-target effects in populations of cells. A number of methods for systematic off-target identification have been developed, each with its own caveats in terms of false positives and false negatives and ability to detect events in therapeutically relevant cellular contexts.\textsuperscript{155} Also the basal rate of DSBs and mutagenesis that exists in normal somatic tissues\textsuperscript{158} including HSCs,\textsuperscript{159} must be taken into account.

In addition to design of a highly unique recognition sequence, several other technical approaches may promote specificity of genome editing, including simply minimizing the duration and degree of expression of the targeted nuclease. ZFNs may be engineered at both their DNA recognition domains and linker and dimerization domains to maximize specificity.\textsuperscript{160} For megaTALs, additional TALE RVD recognition modules may improve effective specificity.\textsuperscript{100} For CRISPR/Cas systems, a number of technical strategies have been explored, including short gRNAs (of 17-18 nt), nickase mutants of Cas9 coupled to paired gRNAs targeting opposing strands, dCas9-FokI dimers for enforced heterodimerization, mutant Cas9 coupled to a programmable DNA binding domain such as ZFP or TALE,\textsuperscript{161} and engineered variants of Cas9 that reduce its interaction strength with the gRNA:target DNA heteroduplex,\textsuperscript{162,163} each of which appear to improve specificity.

For therapeutic translation, these approaches need to balance a favorable portfolio of cellular delivery, on-target efficiency, and off-target specificity, as measured under clinically relevant conditions.

Safety

In contrast to the above molecular considerations, the ultimate utility of therapeutic genome editing approaches will be judged based on clinical safety and efficacy. It is important to note that unlike insertional mutagenesis, which is widely described as a mechanism of tumorigenesis, examples of biologically relevant nuclease off-target genotoxicity under conditions that would approximate therapeutic genome editing are vanishingly scarce. The vast majority of off-target cleavages would be expected to result in neutral indels at noncoding sequences. Therefore, rational nuclease design ought to prioritize avoiding cleavage within critical sequences, such as within recurrently mutated tumor suppressor genes. Extrapolating from regulatory requirements for conventional gene therapy, evaluation for any biological aberrations of target cells, such as propensity for clonal outgrowth or tumorigenesis in vitro or in animals, will justifiably receive emphasis rather than reliance on merely molecular outcomes.

Conclusions

Genome (and epigenome) editing offers the opportunity to prospectively alter DNA (and associated chromatin) to investigate fundamental
genome function and to ameliorate disease. The fast pace of technological advancements in these areas appears poised to accelerate knowledge of gene regulation and genome editing. Multiple preclinical and early stage clinical programs are underway, initially using ZFNs, TALENs, and megaTALs, with CRISPR/Cas-based therapies swiftly following. The continued improvement of the delivery and specificity of these reagents promises to hasten clinical implementation of genuine precision medicine.

Acknowledgments

M.D.H. is supported by the National Institutes of Health (NIH), National Heart, Lung, and Blood Institute grant T32HL007574 and D.E.B. by the NIH, National Institute of Diabetes and Digestive and Kidney Diseases grant K08DK093705 (Career Development Award), the Doris Duke Charitable Foundation, the Charles H. Hood Foundation, the American Society of Hematology, the Borough’s Wellcome Fund, and Cooley’s Anemia Foundation.

References


Authorship

Contribution: M.D.H. and D.E.B. wrote and edited the manuscript.

Conflict-of-interest disclosure: D.E.B. is an inventor on a patent application related to therapeutic genome editing of BCL11A, served as a consultant to Editas Medicine and CRISPR Therapeutics, and received research support from Biogen. M.D.H. declares no competing financial interests.

Correspondence: Daniel E. Bauer, Boston Children’s Hospital, 1 Blackfan Circle, Karp RB 07007A, Boston, MA 02115; e-mail: bauer@bloodgroup.tch.harvard.edu.


A genome editing primer for the hematologist

Megan D. Hoban and Daniel E. Bauer