CRISPR-derived genome editing therapies: progress from bench to bedside

Holly A. Rees[1], Alex C. Minella[1], Cameron A. Burnett[2], Alexis C. Komor[2]*, Nicole M. Gaudelli[1]*

[1]Beam Therapeutics, Cambridge, MA, USA
[2]University of California San Diego, San Diego, CA, USA
*correspondence: ngaudelli@beamtx.com, akomor@ucsd.edu

Abstract

The development of CRISPR-derived genome editing technologies has enabled the precise manipulation of DNA sequences within the human genome. In this review, we discuss the initial development and cellular mechanism of action of CRISPR nucleases and DNA base editors. We then describe factors that must be taken into consideration when developing these tools into therapeutic agents, including the potential for unintended and off-target edits when using these genome editing tools, and methods to characterize these types of edits. We finish by considering specific challenges associated with bringing a CRISPR-based therapy to the clinic: manufacturing, regulatory oversight and considerations for clinical trials that involve genome editing agents.

Introduction

Since the development of molecular cloning, there has been a great deal of scientific interest in altering DNA inside patient cells for therapeutic benefit. Initial attempts focused on gene transfer therapy: insertion of foreign DNA into the genomes of cells, most commonly using a viral vector. This was first attempted to treat severe combined immunodeficiency (SCID) patients in 1990[1], and while clinical efficacy was demonstrated, there were devastating instances of insertional oncogenesis in some studies which were attributed to the retroviral vectors used[2]. However, this same gene transfer therapy strategy has been applied successfully *ex vivo* recently, with favorable assessments of safety and efficacy data. Specifically, five autologous T cell-derived therapies (axicabtagene ciloleucel, lisocabtagene maraleucel, brexucabtagene autoleucel, idecabtagene vicleucel, and tisagenlecleucel), in which T-cells are engineered *ex vivo* via retroviral or lentiviral transduction to express chimeric antigen receptor (CAR) constructs, have been FDA approved for the treatment of patients with B cell malignancies[3-8]. Two AAV-based in vivo gene therapies, in which DNA encoding for a gene of interest is delivered into cells but does not genomically integrate, have also been approved for use in the USA. These include onasemnogene abeparvovec, which delivers the survival motor neuron (*SMN*) gene for the treatment of patients with spinal muscular atrophy type I (SMA)[9,10], and voretigene neparvovec, which delivers the retinal pigment epithelium-specific 65 kDa (*RPE65*) gene for the treatment of patients with Leber congenital amaurosis[11].

In general, non-integrating gene therapy approaches have the potential to be curative for certain conditions, but they are not able to repair a pathogenic mutation in the native context of a patients' own genome. In many instances, including pathogenic dominant mutations, providing a copy of a healthy gene cannot ameliorate symptoms of a disease. Further, gene therapy will not restore endogenous gene expression levels because the delivered foreign DNA is either extrachromosomal or integrated into the genome at a site other than the endogenous locus. For example, a recent mouse study indicated that long-term overexpression of SMN from an AAV9 vector caused aggregation of SMN in the cytoplasm, leading to splicing irregularities and transcriptome-wide RNA dysregulation.[12] In contrast to gene therapy agents, gene editing technologies such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), CRISPR/Cas nucleases (which are currently being developed for various illnesses, including to restore vision loss in LCA10 patients[13] or induce gamma globin production in hemoglobinopathy patients[14]), base editors[15,16], and prime editors[17] modify genomic DNA at the endogenous locus, enabling maintenance of endogenous gene regulation and, in many cases, ablating the pathogenic mutation. There are several on-going efforts to apply gene editing therapies to patients, for example the development of a Cas9-nuclease based approach to restore dystrophin expression in animal models of Duchenne's Muscular Dystrophy[18,19], and the development of a Cas9 nuclease strategy to reduce serum concentrations of transthyretin protein in patients with transthyretin amyloidosis[20]. Here we discuss key challenges and considerations during the progression of CRISPR-based genome editing tools from the bench to the bedside.

Technology inception and* in cellulo *development

Nuclease-based genome editing strategies originated in the 1990's when meganucleases were first used to modify pre-defined endogenous loci in mammalian cells[21]. This major class of genome editing tool works by first introducing a double-stranded DNA break (DSB) at a locus of interest. Subsequent processing of the DSB by the endogenous cellular DNA repair pathways then dictates the outcome. Processing by non-homologous end joining (NHEJ) will result in insertions and deletions of a small number of nucleotides (indels) at the site of the DSB[22,23] (Figure 1). The sequences of these indels are in general reproducible (and can, to certain degrees, be predicted), but cannot be dictated and depend on the local sequence context surrounding the DSB[24-29].

Homology directed repair (HDR) can also process the DSB in the presence of an exogenous supplied piece of DNA that bears homology to the locus as well as the intended edit. HDR will use this as a template to fix the damaged genomic DNA and in the process incorporate the edit into the genome, producing precise, researcher-defined outcomes[30] (Figure 1). Due to these competing pathways, DSB-reliant genome editing often results in a mixture of the desired (template incorporation via HDR) and undesired (indel formation) outcomes. The ability to design and generate nucleases that can introduce a DSB with high efficiency at a custom sequence of interest is of paramount importance to enable the use of these technologies. Due to the extreme difficulty of reprogramming meganucleases to recognize custom sequences, they were quickly replaced by ZFNs and TALENs, which were in turn largely sidelined by the biochemical characterization of CRISPR-Cas9 in 2012[31-36].

It was discovered that the Cas9 enzyme, previously known to be a major player in the interference mechanism of archaeal and bacterial immune systems, could be re-programmed to recognize, bind, and cleave a custom DNA sequence using simple Watson-Crick-Franklin base pairing properties between the locus of interest and a piece of RNA (called the guide RNA, or gRNA) that the enzyme uses for target recognition[31-35] (Figure 1). The target sequence (called the protospacer) also requires the presence of a short protospacer adjacent motif (PAM) for Cas9 binding. The quick turnaround time now achievable between design and validation of new gRNAs (which takes days instead of months, as with ZFNs and TALENs), combined with the high efficiency of DSB introduction by Cas9 in mammalian cells and in animals[35,37] are major factors in the widespread and fast adoption of the technology by a wide swath of researchers and its rapid application in the therapeutics space.

DNA base editors are another class of CRISPR-based genome editing tool. Base editors comprise a catalytically disabled or inactivated Cas9 enzyme[31-35] that does not directly generate DSBs and a deaminase enzyme capable of acting on a nucleobase in single stranded DNA (ssDNA). The Cas9 component of base editors acts to locally denature a pre-defined genomic locus, and the deaminase then performs a chemical deamination reaction on the exocyclic amine of target nucleobases within the single-stranded DNA "bubble" (Figure 2). Specifically, DNA base editors fall into two classes: cytosine base editors (CBEs) convert a C●G base pair into a T●A base pair through a uracil intermediate,[38] and adenine base editors (ABEs) convert an A●T base pair into a G●C base pair through an inosine intermediate (Figure 2)[16]. The resulting uracil or inosine-containing intermediates are thought to be processed by DNA repair pathways distinct from those that process DSBs to introduce their respective point mutations.

When progressing a genome editing candidate from a research tool to a therapy, the biological system in which it is being used will change. Tool development is usually carried out in immortalized mammalian cell lines, as this is time- and resource-efficient. The ultimate target (a specific organ within the body, a certain subtype of cells within the body, *ex vivo* editing of primary cells, etc.) will dictate the intermediary systems in which that the candidate will be tested and optimized during preclinical studies. These intermediary systems can include organoids, human primary cells, and animal models, and can have considerable impacts on the efficiency and specificity of the genome editing, as well as the delivery mode required. Specificity here refers to unintended genome edits that can happen in addition to the desired modification and is discussed in great detail in the next section. Strategies to deliver the editor are largely dictated by the system being used and include viral and nucleic acid-based platforms. Considerations related to therapeutic delivery modes are also discussed in more detail later. Importantly, efficiency may change markedly when moving from immortalized cell lines to *in vivo* work for a variety of reasons. Introduction efficiencies of the desired DSB, uracil, or inosine intermediate may be reduced due to delivery difficulties of the new biological system. Additionally, cell type-dependent differences in expression levels of key DNA repair pathways can also impact genome editing outcomes. Specifically, the HDR pathway that DSB-reliant tools depend on is only active in dividing cells. As a result, indels will typically dominate over precision outcomes in non-dividing cells such as neurons. While the exact DNA repair mechanisms that base editors rely on are currently unknown, several studies have demonstrated efficient base editing in non-dividing cell types[39,40]. This suggests base editors may function via more ubiquitous pathways and has lowered barriers to the clinic.

Preclinical assessment: Interrogation of unintended editing outcomes

The term "unintended edits" encompasses any editing event other than the intended therapeutic edit. However, having an unintended edit may not necessarily be injurious for a cell. For example, the unintended edit can occur in conjunction with the beneficial target mutation and end up being inconsequential. To determine whether unintended edits are consequential, careful consideration of their frequency and biological effect, if any,

is required. Here we focus on unintended edits induced by either base editors or Cas-nuclease technologies. The off-target edits induced by Cas nucleases are likely to be similar to those caused by other types of nuclease (for example TALENs or ZFNs), but these will not be explicitly discussed here.

For Cas-based gene editing technologies, unintended edits can be classified as “guide-dependent” or “guide-independent”. Guide-dependent unintended edits are those that occur only when an editor or nuclease is paired with a particular gRNA, whereas guide-independent edits arise from some functionality of the editing complex unrelated to the gRNA.

Guide-dependent genomic edits

There are two types of unintended guide-dependent genomic edits: those that occur at the on-target locus (i.e. within the intended protospacer/PAM region) and those that occur at off-target loci that typically have sequence similarity to the on-target locus.

i. Guide-dependent genomic edits at the on-target locus

For Cas nucleases, classification of edits at the on-target locus as “unintended” depends heavily on the biological goal of the experiment. Nucleases are used to knock out a gene or disrupt a regulatory sequence via the introduction of indels using NHEJ (Figure 1). In such cases, the therapeutic efficacy of the edit will depend on the sequences of the indels that are formed. In these examples, an indel that leads to an in-frame mutation or that does not have the desired biological effect on a regulatory region would be considered an unintended edit. Other times, a nuclease is paired with a donor DNA template and cellular HDR is used to introduce a defined sequence modification close to the DSB site. In these instances, indel introduction at the DSB site, which usually does not lead to therapeutic benefit, would be considered an unintended edit.

For base editors, there are three main classes of on-target, guide-dependent unintended edits: 1) DNA bystander edits, which are deamination-induced mutations of either adenosine bases (in the case of ABE) or cytidine bases (in the case of CBE or, to a much lesser extent, ABE) that happen to lie within the editing window alongside the target base (Figure 3), 2) indels within the protospacer (Figure 3), and, 3) transversion mutations at the target base (Figure 3).

DNA bystander edits are common and can be easily measured along with the desired editing frequency using techniques such as high-throughput sequencing. DNA bystander edits occur if there is an additional editable base within the ssDNA region that is accessible for deamination (Figure 3). Bystander edits therefore depend on the DNA sequence context surrounding the target nucleotide. In many cases, bystander edits lead to silent mutations in the resulting protein so are thought to be without significant biologic effect[41]. For non-silent bystander mutations, gRNA and/or base editor engineering can reduce bystander editing frequencies while maintaining on-target editing[42]. While there are no known examples of CBEs deaminating adenine bases, ABEs occasionally have very low deamination activity on cytosine residues, which can also be considered as bystander editing (but this is typically <1% sequenced reads)[43,44].

Indels and transversion edits at the target nucleotide are typically minor products of base editing, and CBEs induce these at a higher frequency than ABEs (Figure 3). The mechanism by which indel formation occurs with base editors is not fully known, but it is likely related to cellular processing of the U●G mismatch intermediate (Figure 3). The base excision repair pathway processes U●G mismatches quite efficiently. Initially, the uracil N-glycosylase (UNG) enzyme excises the uracil to produce an abasic site (Figure 2)[45,46]. This abasic site, in conjunction with the DNA nick, may lead to an indel (Figure 3). Additionally, further processing of this abasic site-containing intermediate by error-prone polymerases can introduce unintended transversion mutations (Figure 2). CBEs typically encode an uracil DNA glycosylase inhibitor protein (UGI) to minimize the recognition of the U●G mismatch by cellular UNG and reduce these types of unintended edits (Figure 3)[15]. Consequently, base editor constructs lacking UGI have both higher observed rates of indel formation, as well as transversion edits[46,47]. Cellular recognition of an I●C mismatch (the ABE intermediate) is much less active, so no inhibitor is required. The single nick induced by some base editors may also contribute to indel formation through imperfect cellular repair processes.

ii. Guide-dependent unintended edits at off-target loci

Many Cas protein:gRNA complexes can localize to loci in the genome that possess sequence similarity to the on-target locus (Figure 3)[48]. In the case of nucleases, off-target binding can lead to DSB introduction and eventually indel formation at these off-target loci. The same off-target binding can occur with base editors, but

instead of indel introduction at these off-target loci, deamination of adenosines (in the case of ABEs) or cytidines (in the case of CBEs), if they are present in the editing window at the off-target locus, will occur. For this reason, the guide-dependent off-target editing profile of a base editor paired with a particular gRNA usually comprises a subset of the loci edited when a Cas enzyme is paired with the same gRNA.[15,16,49]

Large-scale genomic rearrangements

In addition to production of indels at an off-target locus, if multiple DSBs are generated simultaneously in a cell (for example when one occurs at the on-target locus and another at an off-target locus), translocations can occur[48]. Large (kilo[50]- or mega-base[51] scale deletions have also been reported with single or multiplexed gRNAs (i.e. multiple, different gRNAs used in the same editing experiment) paired with Cas9, as well as chromothripsis (large rearrangement of chromosomes)[52]. The ability of base editors to cause translocations or other kinds of genomic rearrangement, particularly when multiplexed, is much reduced as compared with Cas9 nuclease, presumably because base editors do not directly generate DSBs[53]. Nevertheless, the potential for base editors to create large scale genomic rearrangements must be carefully evaluated for each individual case.

Guide-independent off-target DNA and RNA modulation

For nuclease editors, the most notable guide-independent off-target editing effect is p53-pathway activation in p53-wild-type cells[54] which has been shown to occur without addition of a gRNA, although more research is required to ascertain the full extent of this activity[54].

For base editors, the guide-independent off-target effects arise from the deaminase enzyme. “Spurious deamination” can occur due to arbitrary interactions between the deaminase and cellular RNA or transiently single stranded regions of cellular DNA (Figure 3). When base editors are highly expressed from strong promoters and delivered as plasmids, both ABEs[55,56] and CBEs[57] can cause a transient elevation to the transcriptome-wide level of A-to-U or C-to-U conversion in cellular RNA. Importantly, this activity is dramatically reduced when mRNA delivery of the base editor is performed instead of plasmid delivery[44], and can be reduced further by incorporation of mutations in the deaminase enzyme[44,55,56,58-60].

Some CBEs have the ability to cause low levels of elevated spurious DNA deamination, causing an elevated incidence of genome-wide C•G to T•A mutations relative to untreated controls (Figure 3)[61,62]. This activity may be caused by either the deaminase enzyme itself, the UGI component of the base editor, or both. UGI alone may cause elevated incidence of C•G to T•A mutations by impairing the repair of either deaminase-induced or spontaneous cytosine deamination that occur naturally. CBE-induced spurious deamination can be reduced by delivering the editor as mRNA[58], and can be reduced to undetectable levels by altering the deaminase enzyme or incorporating mutations that reduce spurious deamination activity[58-60,63]. Thus far, there has been very little evidence that ABEs can induce spurious DNA deamination, and several methods have indicated that earlier generation ABEs do not show elevated levels of spurious deamination[44,62]. While newer generation ABEs (called ABE8’s) have since been engineered with more active deaminases, those that have been evaluated by whole genome sequencing (WGS) assays have not displayed elevated levels of A●T to G●C edits, when mRNA delivery of the editor was used[44].

Careful studies for each new application will be necessary to evaluate the risks of both guide-dependent and guide-independent off-target edits for base editors, as the off-target editing frequency is highly variable, and depends on the gRNA, deaminase enzyme, Cas enzyme, cell type and delivery method used.

Off-target risk assessment for nucleases and base editors

For guide-dependent edits, variation within the human genome (both known and unknown) can create or eliminate potential guide-dependent off target binding sites[64,65]. The computational tools that are available to assess guide-dependent off-target effects are limited by the availability of human genome sequencing data, and laboratory-based *in cellulo* or *in vivo* tools are limited by the genomic sequences of the cells and animals that are used for off-target assessment. Nonetheless, *in silico* off-target prediction based on the protospacer sequence of a gRNA and the PAM sequence of the Cas protein used is a common first step in guide-dependent off-target prediction. Easy-to-use tools have been developed that are able to predict potential genomic off-target loci[66], and, more recently, a variant-aware prediction tool has become available[67]. While *in silico* approaches offer quick insights into the potential for a particular guide RNA to cause off-target editing, they may miss *bona fide* off-target loci, and should be considered as an initial screening method to find potential off-target loci.

In vitro methods for off-target site prediction are more labor-intensive than in silico approaches but are essential to screen for additional off-target loci that could not be predicted from the primary gRNA sequence alone. ONE-Seq[68] is a transformative technique for off-target identification and has enabled a broader, albeit guideRNA-sequence-biased, assessment that uses a library-on-a-chip of synthetic DNA instead of depending on genomic DNA for off-target editing quantification, and ONE-Seq can be applied to a nuclease or a base editor. Additional *in vitro* approaches to off-target identification include approaches that utilize purified genomic DNA incubated with purified genome editing proteins and gRNA (CIRCLE-Seq for nucleases[69], Digenome-Seq for nucleases or CBEs[70,71], and EndoV-Seq for ABEs[72]. These *in vitro* tools all require the genome editing agent to function as a purified protein *in vitro* instead of inside a human cell. Two further *in vitro* approaches involve treatment of living cells with the genome-editing agent and gRNA, followed by isolation of the genomic DNA and subsequent innovative molecular biology steps to screen for edited off-target loci. GUIDE-Seq (for nucleases)[48] and, more recently, Detect-Seq (for CBEs)[73] are considered *in vitro* tools since they do not enable direct identification of the DNA sequence outcome an off target edit, and are typically performed with the non-therapeutically relevant plasmid delivery approach.

Since these *in silico* and *in vitro* approaches to identify potential off-target sites are not performed under therapeutically relevant circumstances, the candidate off-target loci must be independently confirmed or refuted. Experiments performed *in cellulo* or *in vivo* using therapeutically relevant delivery methods are required to assess whether a candidate of-target site identified through these screening approaches is indeed edited in a relevant context.

Although it will likely remain impossible to take into account the entire genome sequence of each intended patient of a gene editing therapy, novel tool development and more extensive genome sequencing efforts within the human population will continue to deepen our understanding of the difficult-to-quantify risk posed to patients due to low frequency, unknown genome variations. Guide-independent edits are inherently stochastic, as they are caused by molecular interaction between the deaminase enzyme and cellular RNA or DNA (in the case of base editors) or by an elevated mutational frequency in the case of nucleases[54]. Methods to detect guide-independent off-target effects have been developed and optimized to maximize their sensitivity but remain limited by the constraints of high throughput sequencing technologies. Endogenous cellular cytosine and adenosine deaminases cause varying levels of background deamination, particularly in the cellular transcriptome. Whole transcriptome or whole genome sequencing (WTS or WGS) can be used to measure whether spurious deamination occurs at a higher frequency amongst treated cells as compared with control cells but cannot determine the full complement of nucleotides within the human genome that may be susceptible to off-target editing by spurious deamination.

It should be noted that the most important aspect of any genome editing agent's off-target profile must be related to the biological consequences of the off-target effects generated by the agent. Detecting the different types of off-target edit that occur above a threshold level, while important, is only the first step. More detailed work must go into assessment of how, if at all, consequential those edits are.

The Path to Clinic: i) Overview of regulatory authority oversight

While the breadth of diseases addressable by genome editing illustrates its potential for transformational impact on human health, drug developers must navigate a complex regulatory framework implemented by the relevant health authorities and supported by national or multi-national laws in order to bring a novel therapeutic to clinical studies in patients. Like for all investigational treatment modalities, a data-intensive package must be prepared and submitted to bring an investigational gene therapy from pre-clinical proof-of-concept to early-phase clinical trials (in the US, this package is the investigational new drug (IND) application and in Europe this package is a clinical trial application (CTA)). INDs in the US are typically classified as either commercial or research. An additional type of application is the individual-patient,-expanded access IND[74]. This mechanism may be used with increasing frequency as more investigators gain expertise in methodologies that enable identification of genome editing approaches with potential to treat patients needing customized therapies for life-threatening genetic diseases for which no standard treatments are available. On the other hand, the resource needs required to successfully negotiate the many steps of the IND or equivalent application process creates challenges for investigators and institutions with limited resources and likely will limit the scalability of clinical applications of genome editing in the near term.

Among the required pre-clinical components that must be satisfied are descriptions of studies establishing biological plausibility for the proposed therapeutic modality, justification for the doses to be tested

in humans, description of the therapeutic's absorption (if applicable) and distribution in a relevant *in vivo* model system, and *in vivo* toxicology studies that inform an appraisal of potential risks to humans. For genome editing applications, assessment of off-target biology is a necessary extension of the pre-clinical toxicology data package. Additionally, extensive documentation of chemistry, manufacturing and control (CMC) information is required. This CMC module must include details of the drug substance and drug product, manufacturing materials (e.g. vectors, cells, reagents, and excipients), all product manufacturing procedures, including a description of the final formulation, product stability and testing for sterility, identity, purity, potency and dose[75].

The mechanistic basis of a particular gene editing therapeutic may invoke a specific set of considerations for its path to clinic. For example, a gene editor selective for a mutation-containing protospacer[76] may require IND-enabling preclinical studies to be completed using reagents that contain the same mutation intended to be targeted in human subjects. Another set of considerations relates to gene editors using a common nuclease or non-nuclease-based editor selected for a rare genetic disease shared by individuals with different pathogenic mutations necessitating one of several potential guide RNAs to be selected corresponding to the mutation position. In this context, whether or not a specific regulatory framework will enable an editor with multiple potential guides to be considered a single investigational product could have significant impact on its clinical development.

The clinical trial protocol(s) within the IND details the elements of studies to be performed in patients, including enrollment criteria, study interventions and overall trial conduct. Key documents supporting the protocol, such as informed consent forms, are included. Summaries of any prior data (safety and efficacy) resulting from use of the investigational therapy in humans are also included. The Investigator's brochure (IB, submitted with the IND summarizes pre-clinical data that are relevant to the study of the investigational therapy in patients as well as prior clinical results (if any) and communicates an initial understanding of potential risks and benefits to investigators who are directly overseeing the care of patients enrolled in the study. This information is required to be updated periodically, according to regulatory authority guidance and when there are important new findings affecting the assessment of risk vs. benefit.

The Path to Clinic: ii) Manufacturing considerations

gRNA optimization

Producing high-quality, therapeutic grade gRNAs at commercial scale is an objective that biotech companies and pharma are actively pursuing. There are at least three challenges related to the length of these oligos that organizations need to conquer when bringing a therapeutic forward that includes a synthetic RNA of this length (>100 nucleobases for *S. aureus* and *S. pyogenes* Cas9) forward: 1. Surmount challenges related to the synthesis of a long nucleic acid polymer, 2. Identify the target oligonucleotide and unwanted side reactions that occur during chemical syntheses through analytical development, and 3. Scale-up manufacturing of these long synthetic gRNAs.

A primary challenge towards creating synthetic sgRNAs for SpCas9 gene editing therapies is due to the long length of the polymer. The gRNA for SpCas9 is typically ~100 nucleotides and are even longer for effectors such as SaCas9, prime editors, and many others. Unfortunately, standard nucleic acid polymerization approaches via solid phase syntheses, developed over the past decade, suffer from poor crude yields (typically 10-25% depending on the sequence and scale) and low overall product purity (post HPLC one can expect to achieve 60-80%) at this length. To date, the vast majority of oligonucleotide therapeutics that have gone to the clinic and marketed are substantially smaller (typically 20-50 nt; e.g. siRNAs, antisense oligonucleotides (ASOs))[77,78] and new synthetic methods must be developed to be compatible for oligonucleotide polymers >50nt in length. Initial solutions, however, ultimately proved to be of little benefit for clinical applications.

One approach towards accessing guide RNA reagent more robustly involved leveraging the natural dual guide system[79].This approach disconnects the single guide RNA into two shorter and more synthetically accessible pieces. However this dual guide RNA approach creates a more complicated drug product overall and gene editing efficiencies are typically lower in eukaryotic cells with a complexed two-component guide versus a single-guide RNA. Another approach involves a truncated guide RNA from the 3' end, however these guides are not suitable for technologies such as prime editors which rely on a user-defined 3' extension of the sgRNA, and typically underperform as they appear to form RNP complexes less robustly than full-length gRNA in the intracellular environment[80,81].

More recently, new chemical approaches have been developed to synthetically access RNA polymers for long, sgRNA synthesis. One such approach generates gRNAs through a convergent, split and ligate synthesis wherein the polymer is synthesized in smaller fragments, and then stitched together. Reported synthetic strategies leverage an azide-alkyne cycloaddition reaction that can robustly “click” together the more accessible smaller RNA fragments into the desired larger single gRNA reagent.[82,83] While this advancement proved to be a step closer towards manufacturing gRNAs of lengths >50 nt with both higher purity and final yield, ligation of the two RNA components generates an artificial triazole linkage that may not be suitable for clinical use without extensive pharmacokinetic/pharmacodynamic (PKPD) experiments to evaluate any toxicities associated with these non-natural linkages.

Other chemical approaches aim to change the process for solid-phase synthesis of RNA itself, which has largely been unchanged since the 1980s, by substituting the standard tertbutyldimethylsilyl (TBDMS) protecting group for a less bulky one. Notably, one of the main challenges towards creating long, sequence defined RNAs using phosphoramidite chemistry utilized in solid phase synthesis protocols, is accessing orthogonally protected monomer building blocks that can be coupled efficiently and in high yields. While the synthetic strategies for DNA have overcome hurdles associated with creating oligonucleotide chains greatly exceeding 100 nts in high yields using solid-phase synthesis, similar outcomes with RNA synthesis have lagged behind. By comparison with DNA, achieving such lengths at high yield with traditional RNA solid-phase synthesis has suffered due to the added complexities associated with the additional 2’OH of RNA which must be protected to inhibit side reactions. These necessary protecting groups can thus cause additional steric clash during coupling of nucleotides which severely limits the length and yield of synthetic RNA. As such, innovation in chemistries have been achieved that replace earlier techniques and utilize protection/deprotection schemes that are optimized for RNA nucleosides that are less bulky and thus more effectively couple, such as 2’-O-thionocarbamte, or cyanoethoxy methyl (and derivatives), to enable simple and robust solid-phase oligonucleotide polymerization.[84,85] Additionally, changing the TBDMS protecting group could also enable more green and less harmful downstream processing steps, such as the use of solutions of hydrofluoric acid used for deprotection of TBDMS groups.

A second key challenge of therapeutic oligonucleotide manufacturing is analysis of the long target RNA, side products and byproducts generated from the synthesis of the materials. For gene editing applications, side products formed during synthesis that deviate from the desired sequence of the sgRNA, can affect gene editing outcomes such as overall efficiency of on target editing and/or formation of additional guide-dependent off-targets. Assays utilizing UV spectrometry, HPLC, UPLC-MS, capillary gel electrophoresis (CGE) all may be developed to fully characterize the quality and purity of sgRNA synthesis, as is typically performed with shorter therapeutic RNAs.[86] High resolution HPLC methods for purity evaluation[87] and methods to sequence target sgRNA must, however, continue to be developed for these longer RNA molecules. For example, while tandem mass spectrometry is typically used to sequence siRNA and ASOs, the length of sgRNA greatly challenges the use of this method. For this reason, to verify the sequence integrity of these longer synthetic RNAs, assays developed that leverage next generation sequencing resolution are being examined to qualify the correct sequence integrity of the strand. For any clinically relevant oligonucleotide utilized for current gene editing therapies, many analytical assays and tools will continue to need to be developed.

The third major challenge is scale-up and manufacturing of these long RNA oligonucleotides required for gene editing. For the first time, organizations need to perform the very difficult synthetic gymnastics outlined above for these longer RNA molecules (>50 nts), on large scales (gram to kilogram) than is required for process development on a research scale. While some chemistries have been developed to improve the overall yield of the polymerization reactions on a small scale, these strategies may not be compatible with large-scale (grams to kilograms) needed for commercialization due to cost and sourcing of reagents and process changes related to larger scale reactions (e.g., longer coupling times). For example, unwanted side products, such as double additions that form duplications in the gRNA sequence, may increase upon scale up leading to a small, but potentially meaningful subset of gRNAs with mismatches (relative to their intended target sequence) in the protospacer. In addition to sequence fidelity, chemists must also be mindful of manufacturing challenges associated with incorporation of modifications commonly used to the modify the RNA backbone such as phosphothioate, 2’OMe, and 2’F. These chemical enhancements are necessary to increase stability of the oligonucleotide in cells[88,89] and are typically incorporated into all synthetic gRNA clinical reagents. Of note, *in vitro* transcription (IVT) solutions, which deploy enzymes to create polynucleic acid linkages, cannot be leveraged for gRNA synthesis due to the need of controlled incorporation of the modified RNA backbone necessary for reagent stability within the cell.

IVT-derived mRNA optimization

Messenger RNA encoding CRISPR-associated Cas9 gene editors as mRNA has emerged as a promising reagent in delivery for both *ex vivo* (electroporation) as well as *in vivo* (lipid nanoparticle (LNP)) delivery applications. mRNAs that encode Cas9-based gene editors including SpCas9 nucleases, base editors, and prime editors range in size from ~4,000 to greater than ~5,000 nucleobases, depending on the editor. To generate RNA of these lengths, *in vitro* transcription (IVT) strategies have been applied. Since the first report in 1990 describing the successful use of IVT-derived mRNA that induced the translation of the corresponding encoded protein in animals, many advancements have been made to enable mRNA to be used as therapeutic tool. Notably, mRNA-LNP formulations has garnered global attention recently as a dominate form of vaccine administration during the COVID-19 pandemic (e.g., Moderna and Pfizer mRNA vaccines). One of the beneficial attributes of mRNA is that protein expression from mRNA is transient, and in the context of gene editing this is a beneficial safety feature helping to avoid off-target effects and immunogenicity that could result from durable gene editor expression.[90] Conveniently, the *in vivo* half-life of mRNA can be controlled via the total exposure time of the drug substance, which is influenced by the dose and delivery modality. Also, LNPs carrying mRNA can be administered repeatedly since no anti-vector immunity is present such as with AAV or other viral technologies.

In order to create a robust mRNA transcript for clinical delivery there are several key material attributes that must be optimized. First, mRNA is typically codon optimized to achieve maximum levels of translation within the cell[91]. Various computer algorithms can be used to generate numerous sequence variants, but the optimal codon usage must be experimentally determined often through *in vitro* screening of series of candidate sequences within the intended cell type. In addition to codon optimization, mRNA reagents are typically modified to enhance stability *in vivo.* Functional enhancements to the mRNA include a 5' cap, 5'- and 3' untranslated region (UTR)[92] and an extended, 3' poly(A) tail[93].

For robust translation of mRNA within a cell, a 5'cap structure is incorporated into the mRNA reagent since natural eukaryotic mRNA contains a 7-methylguanosine (m^7G) modification linked to the mRNA during transcription. Binding of this cap by either translation initiation factors (e.g. EIF4E) or to decapping enzymes (e.g. DCP1) regulates the balance of mRNA translation[94] and decay[95]. To date, there are many different 5' cap analogues that can be utilized including m^7GpppG, anti-reverse cap ($m_2^{7,3,-O}GppG$) and phosphorothioate-modified cap, and researchers chose the most appropriate modification based on cell type and cell differentiation state[96].

Another feature of mRNA transcripts is the inclusion of UTR sequences before and after the encoded protein on the mRNA. Many IVT mRNAs include 3'-UTRs of alpha and beta globin mRNAs that increase stability and some cellular and viral UTRs can also be encoded such as elongation factor 1alpha (EEF1A1[97]) and UTRs from orthopoxvirus[98].

Last, IVT mRNA typically includes a terminating, polyadenylate (poly(A)) tail which further influences the stability and translational efficiency of the mRNA in partnership with the other enhancements on the mRNA transcript. It has been found that the optimal length of the poly(A) tail should be between 120 and 150 nucleotides[99,100].

In addition to optimization considerations for mRNA stability, researchers must also address the inherent immunostimulatory properties of exogenous mRNA delivered for clinical use. Crucially, following IVT of a desired mRNA, the reaction product must be purified from any double-stranded RNA (dsRNA) contaminants produced during the IVT reaction[101]. These dsRNA contaminates are similar to viral genomes and their replication intermediates and thus elicit a type I interferon response, resulting in inhibition of translation and degradation of cellular mRNA and ribosomal RNA[101-103]. Encouragingly, dsRNA contaminants can be efficiently separated from the desired RNA oligomer within the IVT mixture using tailored FPLC or HPLC protocols. Because single-stranded mRNAs, as well as their degradation products, are also known to elicit an immune response when delivered exogenously (endosomal sensors such as Toll-like receptor 7 (TLR7) and (TLR8)[104] are sensitive to ssRNA) additional consideration must be made to evade type I interferon production such as through the incorporation of modified nucleosides such as pseudouridine[105-108], 5-methoxyuridine[109], and 1-methylpseudouridine[110,111].

Concomitant with mRNA reagent optimization, delivery strategies must be developed. For *ex vivo* strategies, conditions are optimized to co-transfect synthetic gRNA and mRNA through electroporation conditions to cells of interest (e.g., CD34+ cells for sickle cell anemia, T-cell engineering) followed by re-infusion

of the transfected cells. For *in vivo* applications, LNP formulations are typically leveraged as they are a robust vehicle for mRNA delivery[112,113]. Of note, most current mRNA-complexed LNPs mainly target the liver when administered systemically due to binding of apolipoprotein E and subsequent uptake by hepatocytes[114]. Current efforts by multiple organizations across the field are actively working to modulate the specificity of LNPs to target other cell types for *in vivo* administration. New strategies toward cell specific LNP formulations may be developed once advancements in fundamental understanding of the mechanisms guiding mRNA escape into the cytoplasm are more fully understood. While outside the scope of this review, great progress has been made in developing delivery vehicles for gene editing payloads involving both mRNA-LNP formulations[115,116] as well as viral strategies including AAV[117,118].

Protein optimization for RNP delivery

An alternative to delivery of a nuclease or base editor as an mRNA is to deliver the genome editing agent as a purified ribonucleoprotein (RNP) complex that is pre-complexed with a gRNA. This can offer the advantage the molar stoichiometry of the genome editing protein and the gRNA can be precisely matched and controlled. Typically, synthetic and chemically protected gRNAs are used as described above as they offer superior potency to unprotected gRNAs[88]. For example, a press release from Editas Medicine[119] demonstrates that RNP delivery of AsCas12a can mediate effective editing in human hematopoietic stem and progenitor cells (HSPCs).

Purification of both Cas nucleases and base editors that have functionality in human cells has been performed at a research scale by electroporation or lipofection[49,120] and AsCas12a nuclease has been purified on cGMP scale[119]. Reports of optimization of Cas9 recombinant expression and subsequent purification outline the process[121,122]. In brief, a pET plasmid vector encoding for inducible expression of the genome editing protein with a purification tag attached is transformed into an *E.Coli* (BL21) strain and the resulting culture grown to mid-log phase. Commonly, a cold shock (4°C incubation) is performed prior to inducing expression of the genome editing protein, and the culture is often grown at a reduced temperature (for example, 18°C) after protein expression is induced[49]. After lysing the resultant E.Coli, the genome editing protein must be purified from contaminating *E.Coli* proteins.

Protein purification from the bacterial vector is typically multi-step, including, for example, a Ni^{2+} column if the purification tag is a His_6, followed by a heparin chromatography column. Further purification steps, for example, a size exclusion column, may further enhance the purity of the genome editing protein, and may enable purification of genome editing proteins without a protein-based purification tag. Protein purity can be assessed using common techniques including SDS-PAGE. It is important that purified proteins contain below an acceptable threshold of bacterial endotoxin[123] (for example, <10 EU/mg protein) and that proteins are stored in an appropriate buffer such that they retain potency over time[121].

Every step of the protein expression and purification process must be optimized for each new construct to maximize protein purity and yield. Applying an existing protein production process to a new construct is often more time consuming than optimizing mRNA production for a new mRNA sequence and yields of recombinant Cas fusion proteins are typically lower than yields of recombinant Cas proteins alone. Nonetheless, RNP delivery remains an active area of research and development.

Viral delivery

Delivery of genome editing agents using viral vectors, most commonly adeno-associated virus (AAV), is an alternative to electroporation or lipid-mediated transfection of mRNA encoding for the genome editing agent plus a gRNA or an RNP complex of the genome editing agent and gRNA. AAV is a common choice for viral delivery because research indicates that AAVs can be non-pathogenic and non-inflammatory, although this must be carefully re-evaluated for each novel application[124]. Further, there are several serotypes of AAV with well-studied tropisms. The packaging capacity of AAV (4.9kbp) can pose challenges for larger genome editing agents, including base editors, that must be split into two separate AAVs. Careful inclusion of a protein intein can enable the base editor to be split into two pieces, with the intein facilitating reconstitution of an intact base editor if the AAVs encoding the two parts are co-infected in a single cell[40,125]. While these split-intein base editors afford efficient editing in mouse models, the manufacturing of two separate AAVs may pose a practical challenge. For nucleases, their smaller size simplifies the use of AAV: for example *S.aureus* Cas9 plus two gRNAs can be encoded into a single AAV; such a construct has been used with AAV serotype 5 to facilitate effective editing in a mouse model of LCA10[13]. AAV production may require optimization for each AAV produced (both from the

standpoint of the sequence encoded in the AAV and the tropism of the AAV). The development and production of effective AAVs, particularly with novel tropisms, remains an active area of research.

Gene editing clinical trials: selected concepts

Excluding numerous studies of engineered, investigational CAR-T and other immune cell types being studied as anti-cancer therapies, over 20 active or pending gene editing trials appear in searchable databases at time of writing (Figure 4). Investigational gene editing therapies are particularly well-suited for addressing rare diseases, as many have monogenic etiologies and no effective, available therapy[126]. Patients with life-threatening illnesses and high unmet clinical need will likely be the focus of enrollment into gene editing clinical trials for the foreseeable future. Similar to investigational anti-cancer therapies, it is not acceptable to enroll healthy volunteers into gene editing trials, due to the attendant toxicity risks and/or potential for permanent sequalae[127]. Unmet need is demonstrated for a particular illness by identifying the treatment gaps remaining for patients in the context of existing therapies approved in a particular regulatory jurisdiction and elucidating how an investigational therapy may fill these. For example, in sickle cell anemia (SCA), other than allogeneic hematopoietic stem cell transplant, which is available to a relatively small proportion of patients, currently approved therapies are neither curative nor do they modify the full range of sources of morbidity and mortality for many patients[128]. Thus, one way current trials of investigational gene therapies in SCA can ensure enrollment of patients with unmet need is to limit eligibility to those with signs and symptoms of poorly controlled disease despite management with existing, approved therapies. Additional eligibility criteria may be implemented to minimize the potential risks of the particular investigational therapy to study patients and to ensure potential clinical benefit can be measured readily, in order to develop an early understanding of the risk/benefit profile.

Rare diseases addressable with an investigational, gene editing approach may require significant effort in advance of trial enrollment to identify eligible patients. Converging efforts to catalog whole genome sequencing data and build rare disease registries, along with partnering with patient advocacy organizations, will be instrumental in enabling diseases to be matched to nascent gene editing strategies as well as patients to available trials[129,130]. Commercial sources of genomic data as well as direct-to-consumer genetic testing may accelerate these efforts. However, sponsors of genome editing studies and investigators must safeguard the privacy of patient-identifying information, including genetic data, and prevent their misuse[131] as part of Good Clinical Practice (GCP), which is an international ethical and quality standard for all aspects of clinical trial design and conduct ensuring integrity of clinical trial data and protection of human subjects.

In addition to study population, the selection of study endpoints is the key driver of clinical trial design, determining the nature and timing of patient-directed assessments and study conduct, database design and statistical analyses. For most early-phase studies, safety and pharmacokinetic/pharmacodynamic endpoints generally are dominant, whereas later phase studies will generally emphasize endpoints that enable refining risk/benefit assessment and understanding therapeutic effect size, often in relationship to a prospectively-evaluated comparator treatment. As a component of GCP, one must consider whether opportunities exist for reducing bias in clinical study designs, including randomizing subjects to treatment arms, using placebo controls, or blinding subjects, investigators and sponsor-affiliated parties to treatment assignments. However, in rare disease trials, these strategies often are impractical or impossible, and treatment efficacy and clinical benefit may be evaluated within a first-in-human study without a comparator group.

Beyond subject and endpoint selection, the clinical study design for investigational gene editing therapies will depend heavily on the delivery modality (e.g. *in vivo* vs. *ex vivo* editing, autologous vs. allogeneic gene edited cell product). For example, dose selection for an *ex vivo*-edited, autologous hematopoietic cell therapy may be based on pragmatic considerations, such as the cell dose capable of reconstituting host hematopoiesis[38], whereas other delivery modalities may warrant a dose escalation design in early-phase clinical testing. Due to the nature of gene editing therapies as well as the possibility that some delivery modalities will not be amenable to repeat dosing, the starting dose in a dose exploration study cannot intentionally be set at a sub-therapeutic level. The ability to dose a particular gene editing therapeutic once versus multiple times is a major determinant of how emerging safety data will be incorporated into clinical decision making within the context of a study protocol. Safety assessments implemented for an investigational gene editing therapy should integrate an understanding of potential risks suggested by the analysis of non-clinical toxicology data, including off-target results, the specific delivery modality, and underlying disease state within the trial design. Sentinel dosing and stopping rules may be employed to minimize risks to additional subjects while the safety of a particular editing

approach is being evaluated in individual subjects. Furthermore, given the potential for permanent change to host cell genomes and delayed adverse events, health authorities including the FDA have required that gene editing clinical protocols include long-term safety observation periods that can last as long as 15 years[132].

Initial determinations of efficacy in early-phase gene editing trials may necessitate a combination of standard clinical assays and pharmacodynamic and biomarker-based approaches that either directly or indirectly reflect the functional outcome of a desired edit. As a particular gene editing therapy moves past initial proof-of-concept testing in humans, execution of larger scale clinical trials to clarify effect size and define risk/benefit profile may be challenging due to scarcity of patients whose disease characteristics match the eligibility criteria. While health authorities may accord some flexibility in study design, particularly for rare diseases[127,133] compared to investigational therapies undergoing late-stage clinical trials for more common illnesses with multiple approved and available therapies, rigor in efficacy assessments will be a requirement for gene editing modalities. Ultimately, efficacy must be related to endpoints that are based on demonstration of clinical benefit (improvement in how a patient feels, functions, or survives) for a therapy to gain marketing authorization, though validated surrogate endpoints may be possible under conditional approval pathways offered by some health authorities.

Access and Equity

It is imperative that careful ethical considerations (in addition to safety measures) be taken as precision genetic medicines are developed from powerful technologies into personalized medicines. One major ethical consideration is to understand how these novel therapies will be paid for by society as whole, given that most healthcare systems and payers are accustomed to complex disease treatments delivered over a lifetime rather than one-time curative therapies. Furthermore, we as a society should ensure that financial status of potential patients does not prohibit access to personalized medicine as this will only exacerbate the already considerable health disparities present in many societies today.[134,135]

Additionally, there are several ethical issues to consider during the process of selecting disease targets. Many practical factors influence target selection, such as assessment of the potential relative risk/benefit ratio for a particular novel therapeutic based on preclinical data. However, we have a moral obligation to also ensure that the populations from which these genetic diseases derive are diverse and not specific to demographics that are already highly privileged. The recent focus on the development of cell therapies for SCA (which disproportionately affects individuals of sub-Sahara African descent) is a good example that demonstrates equity and diversity in disease selection.

Furthermore, there are current inequities in genome-wide association studies (GWAS) that need to be addressed[136,137]. GWAS are the primary mechanism by which we identify genetic disease in the first place. Currently, over 88% of GWAS participants are of European descent[138]. This propagates into an over-representation of European-specific genetic variants in the collection of genetic diseases from which we are selecting our targets. To alleviate this large inherent bias, enormous changes must be made in our pipeline for genetic disease discovery. Finally, as we continue to expand our knowledge of genotype-phenotype relationships there will be intense ethical debates surrounding the labeling of a phenotype as a “disease” versus a “trait”. For example, genetic variants that cause hereditary deafness have been identified and it may be tempting to label deafness as a disease. However, the deaf community is a vibrant cultural and linguistic minority group, and the discrimination that they experience from the majority population is far more detrimental to their well-being than their auditory capacity[139,140]. In the end, we must seek out the expertise of a diverse group of individuals, including scientists, clinicians, bioethicists, policy makers and patient advocacy groups as we embark on the journey from bench to bedside.

Author Contributions
H.A.R. contributed to manuscript conceptualization, writing, and editing. A.C.M. contributed to manuscript writing, editing, and figure design. C.A.B. contributed to manuscript conceptualization and editing. A.C.K. contributed to manuscript conceptualization, writing, editing, and figure design. N.M.G. contributed to manuscript conceptualization, writing, and editing.

Acknowledgements
A.C.K. and C.A.B. acknowledge financial support from the National Science Foundation under award number MCB-2048207 (to A.C.K.).

Conflict of Interest Statement
H.A.R., A.C.M., and N.M.G. are employees of Beam Therapeutics and are shareholders in the company. C.A.B. declares no competing interests. A.C.K. is a member of the SAB of Pairwise Plants, is an equity holder for Pairwise Plants and Beam Therapeutics, and receives royalties from Pairwise Plants, Beam Therapeutics, and Editas Medicine via patents licensed from Harvard University. A.C.K.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

Keywords
CRISPR; Genome editing; Base editing; Therapeutics; Cas9

References

1 Blaese, R. M. *et al.* T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science* **270**, 475-480, doi:10.1126/science.270.5235.475 (1995).

2 Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**, 3132-3142, doi:10.1172/JCI35700 (2008).

3 Neelapu, S. S. *et al.* Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med* **377**, 2531-2544, doi:10.1056/NEJMoa1707447 (2017).

4 Maude, S. L. *et al.* Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N Engl J Med* **378**, 439-448, doi:10.1056/NEJMoa1709866 (2018).

5 Schuster, S. J. *et al.* Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *N Engl J Med* **380**, 45-56, doi:10.1056/NEJMoa1804980 (2019).

6 Wang, M. *et al.* KTE-X19 CAR T-Cell Therapy in Relapsed or Refractory Mantle-Cell Lymphoma. *N Engl J Med* **382**, 1331-1342, doi:10.1056/NEJMoa1914347 (2020).

7 Abramson, J. S. *et al.* Lisocabtagene maraleucel for patients with relapsed or refractory large B-cell lymphomas (TRANSCEND NHL 001): a multicentre seamless design study. *Lancet* **396**, 839-852, doi:10.1016/S0140-6736(20)31366-0 (2020).

8 Munshi, N. C. *et al.* Idecabtagene Vicleucel in Relapsed and Refractory Multiple Myeloma. *N Engl J Med* **384**, 705-716, doi:10.1056/NEJMoa2024850 (2021).

9 Day, J. W. *et al.* Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy in patients with two copies of SMN2 (STR1VE): an open-label, single-arm, multicentre, phase 3 trial. *Lancet Neurol* **20**, 284-293, doi:10.1016/S1474-4422(21)00001-6 (2021).

10 Mendell, J. R. *et al.* Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N Engl J Med* **377**, 1713-1722, doi:10.1056/NEJMoa1706198 (2017).

11 Russell, S. *et al.* Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**, 849-860, doi:10.1016/S0140-6736(17)31868-8 (2017).

12 Van Alstyne, M. *et al.* Gain of toxic function by long-term AAV9-mediated SMN overexpression in the sensorimotor circuit. *Nat Neurosci*, doi:10.1038/s41593-021-00827-3 (2021).

13 Maeder, M. L. *et al.* Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med* **25**, 229-233, doi:10.1038/s41591-018-0327-9 (2019).

14 Frangoul, H., Ho, T. W. & Corbacioglu, S. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia. Reply. *N Engl J Med* **384**, e91, doi:10.1056/NEJMc2103481 (2021).

15 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).

16 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

17 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).

18 Amoasii, L. *et al.* Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* **362**, 86-91, doi:10.1126/science.aau1549 (2018).

19 Amoasii, L. *et al.* Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aan8081 (2017).

20 Gillmore, J. D. *et al.* CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *N Engl J Med* **385**, 493-502, doi:10.1056/NEJMoa2107454 (2021).

21 Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**, 8096-8106, doi:10.1128/mcb.14.12.8096-8106.1994 (1994).
22 Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* **15**, 321-334, doi:10.1038/nrg3686 (2014).
23 Fishman-Lobell, J., Rudin, N. & Haber, J. E. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol Cell Biol* **12**, 1292-1303, doi:10.1128/mcb.12.3.1292-1303.1992 (1992).
24 Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*, doi:10.1038/nbt.4317 (2018).
25 Chakrabarti, A. M. *et al.* Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol Cell* **73**, 699-713 e696, doi:10.1016/j.molcel.2018.11.031 (2019).
26 Delventhal, K., Miller, B., Pruett-Miller, S. M., Modarai, S., Porter, S. N., Regan, M. R., & Sergison, E. Reproducibility of Indel Formation Rates by Comparing GuideRNA Format and Delivery Method. *Journal of Biomolecular Techniques : JBT* **30**, S34–S35. (2019).
27 Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646-651, doi:10.1038/s41586-018-0686-x (2018).
28 Lemos, B. R. *et al.* CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc Natl Acad Sci U S A* **115**, E2040-E2047, doi:10.1073/pnas.1716855115 (2018).
29 Shou, J., Li, J., Liu, Y. & Wu, Q. Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. *Mol Cell* **71**, 498-509 e494, doi:10.1016/j.molcel.2018.06.021 (2018).
30 Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12**, 224-228, doi:10.1016/0168-9525(96)10019-6 (1996).
31 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
32 Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-232, doi:10.1038/nbt.2507 (2013).
33 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, doi:10.7554/eLife.00471 (2013).
34 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
35 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
36 Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome Res* **25**, 1442-1455, doi:10.1101/gr.190124.115 (2015).
37 Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* **32**, 551-553, doi:10.1038/nbt.2884 (2014).
38 Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia. *N Engl J Med* **384**, 252-260, doi:10.1056/NEJMoa2031054 (2021).
39 Yeh, W. H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat Commun* **9**, 2184, doi:10.1038/s41467-018-04580-3 (2018).
40 Levy, J. M. *et al.* Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* **4**, 97-110, doi:10.1038/s41551-019-0501-5 (2020).

41 Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**, 770-788, doi:10.1038/s41576-018-0059-1 (2018).
42 Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376, doi:10.1038/nbt.3803 (2017).
43 Kim, H. S., Jeong, Y. K., Hur, J. K., Kim, J. S. & Bae, S. Adenine base editors catalyze cytosine conversions in human cells. *Nat Biotechnol* **37**, 1145-1148, doi:10.1038/s41587-019-0254-4 (2019).
44 Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* **38**, 892-900, doi:10.1038/s41587-020-0491-6 (2020).
45 Kunz, C., Saito, Y. & Schar, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci* **66**, 1021-1038, doi:10.1007/s00018-009-8739-9 (2009).
46 Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774, doi:10.1126/sciadv.aao4774 (2017).
47 Wang, L. *et al.* Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res* **27**, 1289-1292, doi:10.1038/cr.2017.111 (2017).
48 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
49 Rees, H. A. *et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun* **8**, 15790, doi:10.1038/ncomms15790 (2017).
50 Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765-771, doi:10.1038/nbt.4192 (2018).
51 Cullot, G. *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat Commun* **10**, 1136, doi:10.1038/s41467-019-09006-2 (2019).
52 Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat Genet* **53**, 895-905, doi:10.1038/s41588-021-00838-7 (2021).
53 Webber, B. R. *et al.* Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat Commun* **10**, 5222, doi:10.1038/s41467-019-13007-6 (2019).
54 Enache, O. M. *et al.* Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat Genet* **52**, 662-668, doi:10.1038/s41588-020-0623-4 (2020).
55 Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci Adv* **5**, eaax5717, doi:10.1126/sciadv.aax5717 (2019).
56 Grunewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol* **37**, 1041-1048, doi:10.1038/s41587-019-0236-6 (2019).
57 Grunewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433-437, doi:10.1038/s41586-019-1161-z (2019).
58 Yu, Y. *et al.* Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat Commun* **11**, 2052, doi:10.1038/s41467-020-15887-5 (2020).
59 Zuo, E. *et al.* A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat Methods* **17**, 600-604, doi:10.1038/s41592-020-0832-x (2020).
60 Zhou, C. *et al.* Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275-278, doi:10.1038/s41586-019-1314-0 (2019).
61 Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289-292, doi:10.1126/science.aav9973 (2019).
62 Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292-295, doi:10.1126/science.aaw7166 (2019).

63 Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol* **38**, 620-628, doi:10.1038/s41587-020-0414-6 (2020).
64 Petri, K. e. a. Global-scale CRISPR gene editor specificity profiling by ONE-seq identifies population-specific, variant off-target effects. *Biorxiv*, doi:10.1101/2021.04.05.438458 (2021).
65 Samuele Cancellieri, J. Z., Linda Yingqi Lin, Francesco Masillo, Amy Nguyen, Nicola Bombieri, Felicia Ciuculescu, Varun Katta, Shengdar Q. Tsai, Myriam Armant, Rosalba Giugno, Daniel E. Bauer, Luca Pinello. Human genetic diversity modifies therapeutic gene editing off-target potential. *BioRxiv*, bioRxiv 2021.2005.2020.445054, doi:10.1101/2021.05.20.445054 (2021).
66 Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* **10**, e0124633, doi:10.1371/journal.pone.0124633 (2015).
67 Cancellieri, S., Canver, M. C., Bombieri, N., Giugno, R. & Pinello, L. CRISPRitz: rapid, high-throughput and variant-aware in silico off-target site identification for CRISPR genome editing. *Bioinformatics* **36**, 2001-2008, doi:10.1093/bioinformatics/btz867 (2020).
68 Karl Petri, D. Y. K., Kanae E. Sasaki, Matthew C. Canver, Xiao Wang, Hina Shah, Hyunho Lee, Joy E. Horng, Kendell Clement, Sowmya Iyer, Sara P. Garcia, Jimmy A. Guo, Gregory A. Newby, Luca Pinello, David R. Liu, Martin J. Aryee, Kiran Musunuru, J. Keith Joung, Vikram Pattanayak. Global-scale CRISPR gene editor specificity profiling by ONE-seq identifies population-specific, variant off-target effects. *bioRxiv*, doi:10.1101/2021.04.05.438458 (2021).
69 Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* **14**, 607-614, doi:10.1038/nmeth.4278 (2017).
70 Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol* **35**, 475-480, doi:10.1038/nbt.3852 (2017).
71 Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12**, 237-243, 231 p following 243, doi:10.1038/nmeth.3284 (2015).
72 Liang, P. *et al.* Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat Commun* **10**, 67, doi:10.1038/s41467-018-07988-z (2019).
73 Lei, Z. *et al.* Detect-seq reveals out-of-protospacer editing and target-strand editing by cytosine base editors. *Nat Methods* **18**, 643-651, doi:10.1038/s41592-021-01172-w (2021).
74 Kim, J. *et al.* Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N Engl J Med* **381**, 1644-1652, doi:10.1056/NEJMoa1813279 (2019).
75 Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs); Guidance for Industry, January 2020. https://www.fda.gov/media/113760/download.
76 Chu, S. H. *et al.* Rationally Designed Base Editors for Precise Editing of the Sickle Cell Disease Mutation. *CRISPR J* **4**, 169-177, doi:10.1089/crispr.2020.0144 (2021).
77 Kulkarni, J. A. *et al.* The current landscape of nucleic acid therapeutics. *Nat Nanotechnol* **16**, 630-643, doi:10.1038/s41565-021-00898-0 (2021).
78 Egli, M. & Manoharan, M. Re-Engineering RNA Molecules into Therapeutic Agents. *Acc Chem Res* **52**, 1036-1047, doi:10.1021/acs.accounts.8b00650 (2019).
79 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).
80 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
81 Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**, 333-339, doi:10.1016/j.molcel.2014.09.019 (2014).

82 He, K., Chou, E. T., Begay, S., Anderson, E. M. & van Brabant Smith, A. Conjugation and Evaluation of Triazole-Linked Single Guide RNA for CRISPR-Cas9 Gene Editing. *Chembiochem* **17**, 1809-1812, doi:10.1002/cbic.201600320 (2016).
83 Taemaitree, L., Shivalingam, A., El-Sagheer, A. H. & Brown, T. An artificial triazole backbone linkage provides a split-and-click strategy to bioactive chemically modified CRISPR sgRNA. *Nat Commun* **10**, 1610, doi:10.1038/s41467-019-09600-4 (2019).
84 Dellinger, D. J. *et al.* Streamlined process for the chemical synthesis of RNA using 2'-O-thionocarbamate-protected nucleoside phosphoramidites in the solid phase. *J Am Chem Soc* **133**, 11540-11556, doi:10.1021/ja201561z (2011).
85 Shiba, Y. *et al.* Chemical synthesis of a very long oligoribonucleotide with 2-cyanoethoxymethyl (CEM) as the 2'-O-protecting group: structural identification and biological activity of a synthetic 110mer precursor-microRNA candidate. *Nucleic Acids Res* **35**, 3287-3296, doi:10.1093/nar/gkm202 (2007).
86 Capaldi D, A. K., Brooks D, et al. Quality Aspects of Oligonucleotide Drug Development: Specifications for Active Pharmaceutical Ingredients. *Drug Information Journal* **46**, 611-626 (2012).
87 Kanavarioti, A. HPLC methods for purity evaluation of man-made single-stranded RNAs. *Sci Rep* **9**, 1019, doi:10.1038/s41598-018-37642-z (2019).
88 Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, 985-989, doi:10.1038/nbt.3290 (2015).
89 Finn, J. D. *et al.* A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing. *Cell Rep* **22**, 2227-2235, doi:10.1016/j.celrep.2018.02.014 (2018).
90 Sahin, U., Kariko, K. & Tureci, O. mRNA-based therapeutics--developing a new class of drugs. *Nat Rev Drug Discov* **13**, 759-780, doi:10.1038/nrd4278 (2014).
91 Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol* **22**, 346-353, doi:10.1016/j.tibtech.2004.04.006 (2004).
92 Ross, J. & Sullivan, T. D. Half-lives of beta and gamma globin messenger RNAs and of protein synthetic capacity in cultured human reticulocytes. *Blood* **66**, 1149-1154 (1985).
93 Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* **5**, 2108-2116, doi:10.1101/gad.5.11.2108 (1991).
94 Modrak-Wojcik, A. *et al.* Eukaryotic translation initiation is controlled by cooperativity effects within ternary complexes of 4E-BP1, eIF4E, and the mRNA 5' cap. *FEBS Lett* **587**, 3928-3934, doi:10.1016/j.febslet.2013.10.043 (2013).
95 Li, Y. & Kiledjian, M. Regulation of mRNA decapping. *Wiley Interdiscip Rev RNA* **1**, 253-265, doi:10.1002/wrna.15 (2010).
96 Grudzien-Nogalska, E., Jemielity, J., Kowalska, J., Darzynkiewicz, E. & Rhoads, R. E. Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells. *RNA* **13**, 1745-1755, doi:10.1261/rna.701307 (2007).
97 Zinckgraf JW, S. L. Modulating gene expression using DNA vaccines with different 3'-UTRs influences antibody titer, seroconversion and cytokine profiles. *Vccine* **21**, 1640-1649, doi:10.1016/s0264-410x(02)00740-5 (2003).
98 Bergman, N. *et al.* Lsm proteins bind and stabilize RNAs containing 5' poly(A) tracts. *Nat Struct Mol Biol* **14**, 824-831, doi:10.1038/nsmb1287 (2007).
99 Holtkamp, S. *et al.* Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* **108**, 4009-4017, doi:10.1182/blood-2006-04-015024 (2006).
100 Mockey, M. *et al.* mRNA transfection of dendritic cells: synergistic effect of ARCA mRNA capping with Poly(A) chains in cis and in trans for a high protein expression level. *Biochem Biophys Res Commun* **340**, 1062-1068, doi:10.1016/j.bbrc.2005.12.105 (2006).

101 Kariko, K., Muramatsu, H., Ludwig, J. & Weissman, D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* **39**, e142, doi:10.1093/nar/gkr695 (2011).
102 de Haro, C., Mendez, R. & Santoyo, J. The eIF-2alpha kinases and the control of protein synthesis. *FASEB J* **10**, 1378-1387, doi:10.1096/fasebj.10.12.8903508 (1996).
103 Liang, S. L., Quirk, D. & Zhou, A. RNase L: its biological roles and regulation. *IUBMB Life* **58**, 508-514, doi:10.1080/15216540600838232 (2006).
104 Tanji, H. *et al.* Toll-like receptor 8 senses degradation products of single-stranded RNA. *Nat Struct Mol Biol* **22**, 109-115, doi:10.1038/nsmb.2943 (2015).
105 Kariko, K. *et al.* Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol Ther* **16**, 1833-1840, doi:10.1038/mt.2008.200 (2008).
106 Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162, doi:10.1016/j.cell.2014.08.028 (2014).
107 Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146, doi:10.1038/nature13802 (2014).
108 Anderson, B. R. *et al.* Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res* **38**, 5884-5892, doi:10.1093/nar/gkq347 (2010).
109 Vaidyanathan, S. *et al.* Uridine Depletion and Chemical Modification Increase Cas9 mRNA Activity and Reduce Immunogenicity without HPLC Purification. *Mol Ther Nucleic Acids* **12**, 530-542, doi:10.1016/j.omtn.2018.06.010 (2018).
110 Andries, O. *et al.* N(1)-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *J Control Release* **217**, 337-344, doi:10.1016/j.jconrel.2015.08.051 (2015).
111 Anderson, B. R. *et al.* Nucleoside modifications in RNA limit activation of 2'-5'-oligoadenylate synthetase and increase resistance to cleavage by RNase L. *Nucleic Acids Res* **39**, 9329-9338, doi:10.1093/nar/gkr586 (2011).
112 Geall, A. J. *et al.* Nonviral delivery of self-amplifying RNA vaccines. *Proc Natl Acad Sci U S A* **109**, 14604-14609, doi:10.1073/pnas.1209367109 (2012).
113 Kanasty, R., Dorkin, J. R., Vegas, A. & Anderson, D. Delivery materials for siRNA therapeutics. *Nat Mater* **12**, 967-977, doi:10.1038/nmat3765 (2013).
114 Akinc, A. *et al.* Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Mol Ther* **18**, 1357-1364, doi:10.1038/mt.2010.85 (2010).
115 Kowalski, P. S., Rudra, A., Miao, L. & Anderson, D. G. Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. *Mol Ther* **27**, 710-728, doi:10.1016/j.ymthe.2019.02.012 (2019).
116 Han, X. M., M.J. Nanomaterials for Therapeutic RNA Delivery. *Matter* **3**, 1948-1975 (2020).
117 Xu, C. L., Ruan, M. Z. C., Mahajan, V. B. & Tsang, S. H. Viral Delivery Systems for CRISPR. *Viruses* **11**, doi:10.3390/v11010028 (2019).
118 Yip, B. H. Recent Advances in CRISPR/Cas9 Delivery Strategies. *Biomolecules* **10**, doi:10.3390/biom10060839 (2020).
119 Medicine, E. (Globe Newswire, https://www.globenewswire.com/news-release/2021/06/11/2245743/0/en/Editas-Medicine-Presents-Preclinical-Data-Supporting-the-Initiation-of-the-EDIT-301-Phase-1-2-RUBY-Clinical-Trial-for-the-Treatment-of-Sickle-Cell-Disease-at-the-European-Hematolog.html, 2021).

120 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**, 73-80, doi:10.1038/nbt.3081 (2015).
121 Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **24**, 1216-1224, doi:10.1038/s41591-018-0137-0 (2018).
122 Karvelis, T. *et al.* crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA Biol* **10**, 841-851, doi:10.4161/rna.24203 (2013).
123 Williams, K. L. The Biologics Revolution and Endotoxin Test Concerns. 331-402. , doi:10.1007/978-3-030-17148-3_8 (**2019**).
124 Miao, C. H. *et al.* Nonrandom transduction of recombinant adeno-associated virus vectors in mouse hepatocytes in vivo: cell cycling does not influence hepatocyte transduction. *J Virol* **74**, 3793-3803, doi:10.1128/jvi.74.8.3793-3803.2000 (2000).
125 Villiger, L. *et al.* In vivo cytidine base editing of hepatocytes without detectable off-target mutations in RNA and DNA. *Nat Biomed Eng* **5**, 179-189, doi:10.1038/s41551-020-00671-z (2021).
126 Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179-189, doi:10.1038/s41586-019-1879-7 (2020).
127 Human Gene Therapy for Rare Diseases; Guidance for Industry, January 2020. https://www.fda.gov/media/113807/download.
128 Demirci, S., Uchida, N. & Tisdale, J. F. Gene therapy for sickle cell disease: An update. *Cytotherapy* **20**, 899-910, doi:10.1016/j.jcyt.2018.04.003 (2018).
129 Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96-102, doi:10.1038/s41586-020-2434-2 (2020).
130 Rosales, X. Q. *et al.* The North American mitochondrial disease registry. *J Transl Genet Genom* **4**, 81-90, doi:10.20517/jtgg.2020.12 (2020).
131 Bonomi, L., Huang, Y. & Ohno-Machado, L. Privacy challenges and research opportunities for genomic data sharing. *Nat Genet* **52**, 646-654, doi:10.1038/s41588-020-0651-0 (2020).
132 Long Term Follow-Up After Administration of Human Gene Therapy Products; Guidance for Industry, January 2020. https://www.fda.gov/media/113768/download.
133 Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products; Guidance for Industry, December 2019. https://www.fda.gov/media/133660/download.
134 Gronde, T. V., Uyl-de Groot, C. A. & Pieters, T. Addressing the challenge of high-priced prescription drugs in the era of precision medicine: A systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. *PLoS One* **12**, e0182613, doi:10.1371/journal.pone.0182613 (2017).
135 Bilkey, G. A. *et al.* Optimizing Precision Medicine for Public Health. *Front Public Health* **7**, 42, doi:10.3389/fpubh.2019.00042 (2019).
136 Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161-164, doi:10.1038/538161a (2016).
137 Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26-31, doi:10.1016/j.cell.2019.02.048 (2019).
138 Mills, M. C. & Rahal, C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat Genet* **52**, 242-243, doi:10.1038/s41588-020-0580-y (2020).
139 Pray, J. L. & Jordan, I. K. The deaf community and culture at a crossroads: issues and challenges. *J Soc Work Disabil Rehabil* **9**, 168-193, doi:10.1080/1536710X.2010.493486 (2010).
140 Hladek, G. A. Cochlear implants, the deaf culture, and ethics: a study of disability, informed surrogate consent, and ethnocide. *Monash Bioeth Rev* **21**, 29-44, doi:10.1007/BF03351265 (2002).

Figure Legends

Figure 1. Overview of nuclease-based genome editing. A. Cas9 first complexes with a guide RNA (gRNA), which comprises a ~20 nucleotide long spacer region on the 5' end, and a ~80 nucleotide long backbone on the 3' end. Binding to genomic DNA requires that the sequence of the locus of interest matches that of the spacer region of the gRNA (this region of the DNA is called the protospacer) and has a protospacer adjacent motif (PAM) directly next to the protospacer. B. After the Cas9:gRNA ribonucleoprotein (RNP) complex binds to the genomic locus of interest, two amino acid residues of the Cas9 protein cleave the DNA backbone of either DNA strand to produce a double-stranded DNA break (DSB). C. The resulting DSB can be processed by two different DNA repair pathways, which will produce different outcomes. Non-homologous end joining (NHEJ) introduces non-specific small insertion and deletion (indel) products at the site of the DSB (red outcome). Homology-directed repair (HDR) will use an exogenously-supplied piece of DNA bearing homology to the DSB site as a template for repair, introducing a user-defined change in the genome (blue outcome).

Figure 2. Overview of base editing. A. Cytosine base editors (CBEs) comprise a single stranded DNA (ssDNA) specific cytidine deaminase enzyme shown here fused to a catalytically impaired Cas9 nickase (cas9n). An uracil glycosylase inhibitor (UGI) peptide is also usually included. The gRNA is designed such that a target C●G base pair is within a specific "window" of the protospacer. After the Cas9n:gRNA complex locally denatures the genomic DNA at the protospacer, the cytidine deaminase enzyme converts the target cytidine into a uracil to produce a U●G mismatch, and the Cas9n nicks the DNA backbone of the opposite strand. As uracil has the same base pairing properties as thymidine, replication or repair of this intermediate using uracil as a template (the nick in the DNA backbone helps to bias repair in this direction) will produce an overall C●G to T●A base pair conversion. Efficient uracil excision by the DNA repair enzyme uracil N-glycosylase (UNG) can result in unintended editing outcomes that are outlined in Figure 3. C●G to non-T●A conversions and indels. The UGI component of CBEs acts to suppress this undesired pathway. B. Adenine base editors (ABEs) work similarly but utilize adenosine deamination chemistry to convert a target adenosine base within the protospacer to an inosine. Inosine excision by the cell's endogenous DNA repair machinery is quite inefficient, thus A●T to G●C base pair conversions comprise the overwhelming majority of editing outcomes with ABEs.

Figure 3. Schematic of "unintended edits". A. Guide-dependent unintended edits can occur at the on-target locus or at an off-target locus. Those at the on-target locus include "bystander" editing by base editors (top right). This occurs when additional adenosine (in the case of ABEs) or cytidine (in the case of CBEs) bases are present within the protospacer and become deaminated alongside the intended target adenosine or cytidine base. Additional unintended edits at the on-target locus include indels and transversion edits (left), which can occur via efficient uracil excision by the DNA repair enzyme uracil N-glycosylase (UNG), which will produce an abasic site. Further processing of this abasic site intermediate can result in an unintended byproduct of base editing: C●G to non-T●A conversions and indels. The UGI component of CBEs acts to suppress this undesired pathway. Unintended edits at off-target loci (bottom right) occur when the Cas9:gRNA complex binds to sites within the genome that are similar to the spacer sequence but do not match completely. In this case, mismatches or bulges between the spacer and the genomic DNA will be present, but Cas9 can still bind and localize long enough for Cas9 to introduce a DSB (or for BEs to deaminate adenosines or cytidines that are present within the protospacer) at these off-target loci. B. Guide-independent unintended edits by BEs occur via "spurious deamination" mechanisms. In these cases, if cytidines (in the case of CBEs) or adenosines (in the case of ABEs) are present in ssRNA or ssDNA within the cell, they can become deaminated.

Figure 4. Active and upcoming clinical trials of gene editing therapies. These are grouped by disease category (Left) and modality (Right), excluding engineered immune effector cell therapies for cancer. Included are active or not-yet recruiting trials listed in Trialtrove (Informa) or clinicaltrials.gov databases, identified using one or more of the following search terms: gene edit(-ing/-ed), genome edit(-ing/-ed), CRISPR, TALEN, zinc finger nuclease (ZFN). Terminated, withdrawn, completed, unknown status, or not-yet-recruiting trials without status updates provided within the last 2 years were excluded. Investigational immune effector cell therapies for cancer, including CAR-T, were not included, due to lack of consistent or clear designation of the editing technology utilized for manufacturing in some cases. Long-term follow-up or extension trials or trials of investigational diagnostics were not included. Values shown are absolute numbers per category. Geographic regions

represented: USA only (13 trials), China only (6) and multi-national (3). HSPC – hematopoietic stem and progenitor cells, CC9 – CRISPR/Cas9