A Beginner's Guide to Gene Editing

Patrick T Harrison
Department of Physiology, BioSciences Institute, University College Cork, Cork, Ireland

Stephen Hart
UCL Great Ormond Street Institute of Child Health, University College London, United Kingdom

---

New Findings

This review summarizes the development of gene editing from early proof-of-concept studies in the 1980's to contemporary programmable and RNA-guided nucleases which enable rapid and precise alteration of DNA sequences of almost any living cell. With an average of one CRISPR Cas9 paper published every four hours in 2017, this review cannot highlight all new developments, but a number of key improvements including increases in efficiency, a range of new options to reduce off-target effects and plans for CRISPR to enter clinical trials in 2018 are discussed.

Abstract

Genome editing enables precise changes to be made in the genome of living cells. The technique was originally developed in the 1980's but largely limited to use in mice. The discovery that a targeted double stranded break (DSB) at a unique site in the genome, close to the site to be changed, could substantially increase the efficiency of editing raised the possibility of using the technique in a broader range of animal models and potentially human cells. But the challenge was to identify reagents that could create targeted breaks at a unique genomic location with minimal off-target effects. In 2005, the demonstration that programmable zinc finger nucleases (ZFNs) could perform this task, led to a number of proof-of-concept studies, but a limitation was the ease with which effective ZFNs could be produced. In 2009, the development of TAL-effector nucleases (TALENs) increased the specificity of gene editing and the ease of design and production. However, it wasn't until 2013 and the development of the CRISPR Cas9/guideRNA that gene editing became a research tool that any lab could use.

Author Contributions

P.T.H. and S.H. designed and conceived the manuscript.
P.T.H. drafted the manuscript.
P.T.H. and S.H. revised the manuscript critically for important intellectual content, approved the final version of the manucript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

Introduction

Genome editing is a technique that enables precise changes to be made in the genome of a living cell. It can be performed in ex vivo culture and selection applied to enrich for edited cells. It can also be performed *ante natal* in essentially any animal species, and has been used to edit single cell embryos, which when implanted into surrogate mothers, develop into adults in which all cells are edited. Early forms of this technique revolutionised the understanding of murine physiology in the 1980's, but limitations in the technology restricted its application to human cells and therapeutic development.

At the dawn of the new millennium, many of these limitations have been resolved and a broad range of gene editing tools are available to precisely target unique sites in the genome. It is now possible to precisely, efficiently and permanently edit the genome of cells from virtually any organism on the planet in order to study complex physiological processes and genetic disorders. Gene editing has been used to establish proof-of-concept as a treatment of such genetic disorders in these model systems, and in

December 2017, Brian Madeau, a 44-year old with Hunter syndrome became the first person to reveive gene editing technology in a phase I/II clinical trials. Moreover, gene-edited cells have already been used successfully in a small number of therapeutic studies in human, and as discussed below, one type of edited cells has recently been approved by the FDA as a licenced medicine.

The purpose of this review is to introduce the basic techniques used to edit (see Box 1) the genome of living cells in the context of three major biomedical applications:

- Establishing how particular DNA sequences control the normal physiology of individual cells in ex vivo culture.
- Exploring how certain mutations in the DNA sequence, either acquired or inherited, can cause more than 100 different types of cancer and over 7,000 genetic disorders.
- Development as the basis for a potential treatment for many of these diseases, such as cystic fibrosis (CF), sickle-cell anaemia (SCA) and B-cell Acute Lymphoblastic Leukaemia/Lymphoma (B-ALL).

INSERT BOX 1

DNA structure, replication and repair

To fully understand editing, we need to understand some of the key features of the structure of DNA, and the mechanisms by which it is replicated and repaired in cells. DNA is made from just four nucleotides (or bases), abbreviated A, C, G and T, arranged as two antiparallel strands in a double helical configuration (Figure 1a). The two strands are held together by hydrogen bonds between the nucleotides which always occur as either an A-T base-pair, or G-C base-pair. As noted by Watson and Crick in their 1953 paper on the structure of DNA, this specific base pairing "*immediately suggests a possible copying mechanism for the genetic material*". If it's not *immediately* obvious, the point they were making is that the anti-parallel arrangement of DNA nucleotides means that if the strands of DNA are pulled apart (unwound), then the two individual strands have all the information necessary to make two new double stranded DNA molecules, provided the two strands are surrounded by the four DNA bases in the trinucleotide form with the appropriate enzymes and cellular environment. Suitable environments range from the nucleus of a mammalian cell to the isolated components in a polymerase chain reaction (PCR) in a test tube.

This understanding of DNA replication underpins an understanding of DNA repair mechanisms, which is the basis for all DNA editing. When the strands separate (Figure 1b), short "primer" molecules bind in a base-pair specific manner (Figure 1c) and initiate DNA synthesis in a 5' to 3' direction. In the upper strand in the diagram, the template strand is undamaged and DNA synthesis continues in an uninterrupted manner (Figure 1d). However, the lower strand in this diagram has been damaged such that the phosphodiester linkage between two of the bases has been broken (Figure 1e) and DNA synthesis cannot cross what is effectively a double-stranded break (DSB). Such DSBs occur routinely during DNA replication but cells have evolved with DNA repair enzymes that are expressed during the synthesis (S) phase of the cell cycle which can repair these DSBs by a process known as homology directed repair (HDR). Repair is initiated by enzymes which recognise the DSB then resect some of the DNA (~300 bases) to expose DNA sequences with free 3' ends. These 3' ends can invade and bind to the newly synthesised sister chromatid in a base-pair specific manner (Figure 1f), and use it as a template for DNA synthesis which then continues across the DSB (Figure 1g). Once the DSB has been crossed, the strands reanneal to the distal side of the DSB and synthesise new DNA to fill in the gaps and complete the HDR process (Figure 1h). As shown below, the HDR repair pathway is key to most precision gene editing strategies. However, it is worth noting that HDR is not the default repair for a DSB, with the most likely outcome for repair of a DSB being non-homologous end joining (NHEJ). This typically results in the joining of the two free ends of DNA with the creation of small insertions of deletions (indels) following repair. Whilst the formation of these indels can be mutagenic, in that gene sequences are altered, the re-joining of the DNA rescues the chromosome and thus enables the cell to replicate. The consequences of NHEJ for precision gene editing strategies are discussed below.

INSERT FIGURE 1

Editing and Genetic Disorders

The concept of genetic disease or inborn metabolic disorders was defined by Archibald Garrod in 1902, but it was another 55 years before Vernon Ingram confirmed that changes in the DNA sequence gave rise to changes in a protein sequence (Ingram, 1957). This work which identified the single nucleotide change in the beta-globin gene (an A replaced by a T) accounted for the replacement of the amino acid glutamine at the sixth position of the haemoglobin B chain with a valine which in turn gives rise to sickle cell anaemia when both copies of the gene carry this mutation which affects millions of people worldwide.

The importance of this finding should not be underestimated as it not only confirmed the hypothesis that a substitution in the DNA leads to a substitution in the amino acid sequence of a protein, but it also underpinned efforts of Holley, Khorana and Nirenberg to successfully crack the genetic code and win the Nobel Prize in Physiology or Medicine in 1968 (www.nobelprize.org/nobel_prizes/medicine/laureates/1968), and the concept of the central dogma of molecular biology. Together, these findings led to the realisation that if techniques could be developed to manipulate DNA sequences in living cells, this would significantly increase our ability to understand the physiological process of that cell. And if that cell was an embryonic stem (ES) cell, then this gives rise to the possibility of studying the effects of manipulated DNA sequences in whole animals derived from that cell.

Of Mice but not Men

Over the next 20 years, the advances in molecular genetics and development of recombinant DNA technology culminated in the development of a gene editing strategy (then known as gene targeting) that enabled precise changes to be made in mouse ES cells. The technique relied upon the generation of long DNA molecules (≥10kb) known as targeting constructs that could be microinjected into ES cells. These constructs contained long homology arms (DNA sequences that precisely matched the DNA sequence of the target gene) flanking the desired DNA sequence to be introduced, as well as positive and negative selection markers. The positive selection allowed the identification (and selection) of cells that had been modified (precisely or randomly), and the negative selection allowed the subsequent identification (and removal) of cells that underwent random targeting. The end result is a small population of precisely targeted (edited) ES cells. The parallel development of protocols to generate mice from ES cells enabled the generation of the first gene edited animal in 1982. This strategy revolutionised the study of mouse physiology and *in vivo* modelling of genetic and physiological disorders, culminating in the 2007 Nobel Prize in Medicine or Physiology (www.nobelprize.org/nobel_prizes/medicine/laureates/2007).

The technique was largely restricted to engineering of mouse ES cells due to the low efficiency and consequent need for strong selection strategies as described above, even though the key to significantly improving the efficiency, **a double-stranded break close to the region to be edited**, was identified in 1984. Oliver Smithies and colleagues established a two-plasmid system to test the idea using restriction enzymes. One of the plasmids contained a reporter gene with a small deletion such that it lost its function, the second contain the same gene with a different deletion which also led to loss of function. When they transfected either plasmid individually into cells, they saw no reporter activity. However, when co-transfected, they detected a low incidence of recombination events where one plasmid acted as a donor to repair the other by HDR. If they made a targeted DSB in one of the plasmids before they did the co-transfection, they doubled the efficiency of HDR. However, if the DSB was made very close to the target site, the efficiency of HDR increased by 22-fold (Kucherlapati et al., 1984). But why wasn't this approach pursued for editing genomes?

Breaking up is hard to do; do the maths!

With apologies to Neil Sedaka, creating a unique double-stranded break in genomic DNA isn't that easy. The reason is simply that unique sequences in DNA must be at least 16 bases long, and until the turn of the 21st century reagents to specifically recognises and cut such a long sequence weren't available (see box 2).

INSERT BOX 2

Programmable Zinc Finger Nucleases – the breakthrough moment for Gene Editing

Zinc finger nucleases are synthetic enzymes comprising three (or more) zinc finger domains, each of which binds three bp of DNA, linked together to create an artificial DNA binding protein binding ≥9 bp of DNA. In order to cut DNA, the zinc finger domains are fused to one half of the *Fok*I nuclease domain such that when two ZFNs bind the two unique 9 bp sites, separated by a suitable spacer, they can cut within the spacer to make a double stranded break. The original creator of this approach used naturally occurring ZF domains assembled to recognise a unique sequence in the DNA genome of the *E. coli* bacteriophage lambda. Subsequent studies by academic and industry groups led to the identification of a set of rules to allow ZFNs to be designed to target essentially any target site in the human genome culminating in precision repair of mutations in the IL-2 receptor γ chain gene by HDR in ~20% of transfected cells (Urnov et al., 2005). This breakthrough quickly led to a number of pivotal proof-of-concept experiments in gene editing including:

- ability to make gene targeted animals by editing in vitro fertilised embryos rather than ES cells (Guerts et al., 2009); prior to this, editing cells from animals other than mice was very time consuming due to lack of ES cell lines. This approach of *ex vivo* editing has been developed in subsequent years to model human disease and study physiological processes in rodents (Birling et al. 2017) and many other species (Whitelaw et al., 2016; Morales and Wingert, 2017).
- ability to edit *in vivo*, correcting multiple mutations at once, and restoring normal physiological function in a disease model (Li et al., 2011); a truly revolutionary study showing the power of gene editing to permanently correct a defect in a mouse model haemophilia B. Moreover, by using a "superexon" (or partial cDNA) as donor, this approach can potentially correct many different disease-causing mutations in the same gene, which is potentially critical for the widespread clinical development of gene editing. Indeed, the first direct *in vivo* use of ZFNs in humans involved a cDNA integration strategy (Sangamo, 2017)

Gene Edited Cells for Clinical Application

A few years after the development of ZFNs, a second gene editing system was reported, the Tal-effector nucleases (TALENs). These reagents have been used by many labs and are now easy to assemble and use for numerous applications. One of the most exciting developments has been use in the development of genetically engineered T-cells which express a chimeric antigen receptor (CAR) which recognises a specific antigen on cancer cells in B-cell Acute Lymphoblastic Leukaemia. The CAR was added to the T cells by conventional genetic engineering approaches, but two additional genetic changes to the T cells to increase the safety and efficacy where introduced by TALEN gene editing. The use of CAR-T cells to treat B-ALL was approved by the US FDA on 30th August 2017 (www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm574058.htm).

The CRISPR Rising: where were you on 17th August 2012?

In spite of the enormous potential applications for both ZFNs and TALENs, the prospect of designing and synthesising these recombinant nucleases to target a gene of interest limited the widespread uptake of these reagents; the field was on the lookout for a simpler way to make a unique cut in the genome of a living cell. Whilst maybe not (yet) as memorable as 20th July 1969 (first human on the moon), 9th November 1989 (fall of the Berlin wall), or 5th July 1996 (birth of Dolly the sheep), 28th June 2012 marked the online publication that the clustered regularly interspaced short palindromic repeat (CRISPR) sequences in the genome of *Streptococcus pyogenes* encoded a dual-RNA structure (tracrRNA and crRNA) that directs a Cas9 endonuclease to make targeted DSBs in DNA, and that this *"could offer considerable potential for gene-targeting and genome-editing applications"* (Jinek et al., 2012). Subsequent studies described modifications to Cas9 such as codon-optimised to maximise expression in mammalian cells, and addition of a nuclear localisation signal to facilitate Cas9 entry into the nucleus. Many studies also utilised a single guide RNA

(sgRNA, often abbreviated further to gRNA) of ~120 bases (Figure 2A), which is essentially a fusion of the tracrRNA and crRNA molecules.

Once expressed in or introduced to the mammalian cell, the Cas9/gRNA complex migrates to the nucleus and scans the genome for a short DNA triplet sequence (normally 5'-NGG-3') known as a PAM. When the Cas9/gRNA binds to a PAM, it then partially unwinds the DNA upstream of the PAM; if the target region of the gRNA can bind with a 20:20 match, the Cas9 nuclease then cuts both strands of DNA 3 bp upstream of the PAM (Sternberg et al., 2014). As predicted above, once Cas9/gRNA has made a unique DSB, then the same DNA repair pathways exploited by ZFNs and TALENs can be used for editing.

INSERT FIGURE 2

Design, Cloning, Testing, Selection and Off-Target

This section gives a brief overview of how to get started with CRISPR in the lab, with more detailed information available in protocol and methodology papers (Ran et al., 2013; Byrne et al., 2014; Wang G et al., 2017).

gRNA target selection – there are numerous freely available online tools which in essence follow three basic rules. First, they identify a PAM sequence in your target gene. Second, they identify the 20-nucleotide target sequence upstream of the PAM. Third, they scan the genome of the species you are intending to edit and assess for potential off-target sequences, that is sequences that occur elsewhere in the genome which match or nearly match your target. Some programmes will also generate the two oligo sequences you need to synthesise to clone into the Cas9/gRNA dual expression vector (see Figure 2B). See Lee et al., 2017 for a review of different gRNA design tools.

Cloning gRNAs – a number of plasmid-based systems are available to co-express Cas9 and a specific gRNA. Once a gRNA target has been identified, the target sequence can be made as a double-stranded fragment of DNA by annealing two short oligonucleotides and using Golden Gate cloning to insert them into a dual gRNA/NLS-Cas9 expression vector. Golden Gate cloning is possibly one of the simplest cloning techniques since all the reagents are added to the same tube then subjected to 10 cycles of 37˚C/16˚C over a 2-3 hr period then transformed into *E. coli*. Typically, ≥80% of the clones isolated the following day contain the correct gRNA cloned in the correct orientation (Ran et al., 2017).

Avoiding cloning altogether – to simplify matters further, cloning can be completely avoided by using recombinant Cas9 protein mixed with in vitro transcribed gRNA molecules (Lingeman et al,. 2017), or chemically modified gRNA which can improve stability and enhance editing efficiency (Hendel et al., 2015). Alternatively, an RNA only approach is possible using the mRNA which encodes Cas9 co-transfected with the gRNA molecule (Wang HX et al., 2017).

Testing gRNAs – if a gRNA is successful at creating a DSB, the cells will normally repair this by NHEJ which usually creates small insertions or deletions (indels; Hollywood et al., 2016). As the indels are usually small (≤25 bp) they are not that easy to detect by conventional PCR. However, if PCR is performed and the final reaction is heated to 95˚C and allow to cool slowly, a significant proportion of heteroduplex molecules should form and these can easily be detected using nuclease which can cut these heteroduplex regions (see Lee et al., 2012). Alternatively, pairs of gRNAs can be tested simultaneously to generate deletions of defined size which can be readily detected by conventional PCR (see Sanz et al., 2017).

Donor Design – as shown in Figure 3 below, the donor molecule is key to introducing the correct sequence once a targeted DSB has been made. The donor can be a single stranded oligonucleotide (c. 100 bases) or a much longer double-stranded DNA molecule usually in the form of a plasmid (Byrne et al., 2014). In both cases, the donor should have homology arms (HA_{left} and HA_{right} in Figure 3), extended regions of sequence identity either side of the region to be edited. The donor should obviously contain the desired nucleotide changes required to introduce or correct mutations in the genome, and it is also possible to make changes

in the donor such that the PAM is removed from the genome to prevent subsequent cutting once the desired change in the genome has been made (see Figure 3).

Reducing off-target effects – one of the early challenges with Cas9/gRNA editing was the relatively high level of off-target effects as discussed in Fu et al., 2013. An off-target effect occurs when Cas9/gRNA creates a DSB at a location within the genome which has a very similar sequences to the original 23 base-pair target site. Such DSBs are typically repaired by the NHEJ pathway which usually results in indel formation at the off-target site which can then have undesired consequences on gene expression and/or function. Numerous options have been devised to prevent off-target effects, possibly the simplest is to make gRNAs 1-2 bases shorter at the 5' end; this can significantly reduce DSB formation at off-target sites whilst retaining sufficient activity at the correct target site (Fu et al., 2014). The use of Cas9/gRNA protein-RNA particles (Kim 2014), and/or Cas9 variants such as Hypa-Cas9 (Chen et al., 2017) have been shown to reduce off-target effects with very little impact on on-target activity,

Selection of edited cells – making precise alterations in the genome of cells is now relatively easy, but the isolation and selection of those isn't necessarily as simple. Numerous protocols have been devised using selection markers which can be subsequently removed (Wang et al., 2017) or PCR-based enrichment strategies to avoid the need for selectable markers (Suzuki et al., 2016).

Precision repair of cystic fibrosis-causing mutations by HDR

One of the major goals of our research is to apply gene editing techniques to better understand and potentially treat Cystic Fibrosis, which is caused by mutations in the *CFTR* gene (Rommens et al., 1989; Riordan et al., 1989; Kerem et al., 1989). To date, ZFNs, TALENs and Cas9/gRNA have all been used to edit CF-causing mutations in a range of cell types using NHEJ and HDR methods (Harrison et al., 2016; Alton et al., 2016; Sanz et al., 2017; Hart and Harrison, 2017). An overview of the mechanism by which Cas9/gRNA HDR is could be used to repair a stop codon mutation in the *CFTR* gene is shown in Figure 3.

INSERT FIGURE 3

Summary

Gene editing is a long-established technique dating back ~30 years BC (before CRISPR), but for much of this time was limited mainly to mouse ES cells and specialised research groups. The development of ZFNs and TALENs set the scene for a revolution, but it took the advent Cas9/gRNA to make gene editing truly accessible to any research laboratory.

It took less than 50 years from the discovery of structure and function of DNA (Watson and Crick, 1953) to the first draft of the human genome sequence (Lander et al., 2001). The impact of these discoveries has already revolutionised our understanding of cell and organismal physiology, and genetic disease and cancer.

It is only year 6 CE (CRISPR Era), but already Genome-Read projects such as 100,000 genomes look set to be eclipsed by Genome-Write projects, and the unstoppable growth in Cas9 editing technology and applications (Figure 4) will undoubtedly play a guiding role. Further advances in technology to re-write the genome should be paralleled by a commitment to both scientific and public discussion on the ethical and regulatory issues surrounding how these techniques will be used (Boeke et al., 2016; Ormond et al., 2017; Parry, 2017, Harrison et al., 2017), and where we might be by 50 CE?

INSERT FIGURE 4

Figure and Box Legends

Box 1: How do you precisely edit a DNA sequence, in a live cell?

Editing a DNA sequence is somewhat similar to editing a text document on a computer.

Step 1 is to find (ctrl F) the precise DNA sequence to be changed then to cut it out (ctrl X). With CRISPR gene editing, a guide RNA molecule in combination with the Cas9 nuclease is used to find the unique target sequence in the genome and cut the DNA at that point.

Step 2 in text editing is to copy (ctrl C) a few letters or a word from a dictionary and paste (ctrl V) them into the correct place to change the meaning of a sentence (and potential the overall outcome of the document). In cells, we don't use a dictionary, rather a synthetic piece of DNA with the desired sequence is inserted into the target cells. The synthetic DNA contains the new DNA sequences as well as some sequence similarity (or homology) to the target region of the selected gene which helps it to be correctly copied and pasted (or ligated) into the cut site generated by Cas9/guide RNA. The end result is genomic DNA with a sequence corresponding to that in the synthetic DNA (and potential alterations to the cell's physiology). In summary, the Cas9/guide RNA makes a targeted break in the genome, and the synthetic DNA provides the new information. The process is mediated in cells by a naturally occurring DNA repair mechanisms known as homology-directed repair (HDR).

Figure 1: Overview of DNA replication and repair of double stranded breaks (DSBs); see text for details.

Box 2: Genomic Maths

DNA has four bases (A, C, G and T), so a 6 base sequence such as GAATTC should occur, on average every 4^6 = 4,096 bases of a random DNA sequence. Enzymes such as the *Eco*RI endonuclease can recognise and cut such 6 base sequences, which is fine if you are working with plasmids (typically ~5,000 bp). However, for the human genome, 3x109 bp, such enzymes would cut ≥750,000 times!

The mathematics is simple: $4^{16} = 4.3 \times 10^9$, showing that a unique site should occur if the DNA nuclease recognises at least 16 bases. As shown below, the three most common editing nucleases have recognition sequences in excess of this number - ZFNs recognise 18-24 bp, Cas9/gRNA recognises 19-20 bp and TALENs recognise ≥24 bp.

Figure 2: Cas9/gRNA. A) Cartoon representation of Cas9/gRNA showing PAM site, target sequence and cut site (red triangles). B) DNA insert encoding target sequence to insert into dual gRNA/ NLS-Cas9 expression vector using Golden Gate cloning.

Figure 3: **CRISPR Cas9/gRNA HDR to correct W1282X mutation in CFTR gene.**

Six key stages are shown in the diagram, using an alternative set of words based on the CRISPR acronym (Jansen et al. 2002); note that Figure 5 is shown in the same format as Figure 1 which explained the HDR repair mechanism. The upper sequence (blue) represents the donor, the lower sequence (black) is the genome with amino acid sequence (three letter code) shown above the DNA sequence and * is the stop codon. Bars either side represent extended stretches of DNA (typically 500-1000 bp for donor, or remainder of chromosome arms for genome)

Cut	Cas9/gRNA makes a DSB 3 bp upstream of the PAM site
Resect	Cellular exonucleases resect the free 5' ends of DNA exposing 3' ends.
Invade	A free 3' end invades donor and binds in a sequence specific manner…
Synthesis	… which triggers DNA synthesis in 5'-3' direction copying the donor sequence
Proof-read	Mismatches are corrected by removing part of one strand …
Repair	… then the gap is filled in. The TGA codon is edited to TGG

Note the donor is deliberately modified in this example to edit the PAM sequence in the genome to prevent recutting by Cas9/gRNA whilst preserving the same amino acid coding capacity.

Figure 4. Publications per year using PubMed search term "CRISPR Cas9".

References

Alton EW, Boyd AC, Davies JC, Gill DR, Griesenbach U, Harrison PT, Henig N, Higgins T, Hyde SC, Innes JA, Korman MS. **Genetic medicines for CF: Hype versus reality.** Pediatr Pulmonol. 2016 Oct;51(S44):S5-S17. doi: 10.1002/ppul.23543.

Birling, M-C, Herault, Y, Pavlovic, G. **Modeling human disease in rodents by CRISPR/Cas9 genome editing.** Mamm Genome. 2017; 28(7): 291–301.

Boeke JD, Church G, Hessel A, Kelley NJ, Arkin A, Cai Y, Carlson R, Chakravarti A, Cornish VW, Holt L, Isaacs FJ, Kuiken T, Lajoie M, Lessor T, Lunshof J, Maurano MT, Mitchell LA, Rine J, Rosser S, Sanjana NE, Silver PA, Valle D, Wang H, Way JC, Yang L. **Genome Engineering. The Genome Project-Write.** Science. 2016 Jul 8;353(6295):126-7. doi: 10.1126/science.aaf6850.

Byrne SM, Mali P, Church GM. **Genome editing in human stem cells.** Methods Enzymol. 2014;546:119-38. doi: 10.1016/B978-0-12-801185-0.00006-4.

Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA. **Enhanced proofreading governs CRISPR-Cas9 targeting accuracy.** Nature. 2017;550:407-10.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. **High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.** Nat Biotechnol. 2013 Sep;31(9):822-6. doi: 10.1038/nbt.2623.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. **Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.** Nat Biotechnol. 2014 Mar;32(3):279-284. doi: 10.1038/nbt.2808. Epub 2014 Jan 26.

Garrod, AE. **The Incidence of Alkaptonuria: A Study in Chemical Individuality**. Lancet 1902, vol. ii, pp. 1616-1620.

Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Ménoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R. **Knockout rats via embryo microinjection of zinc-finger nucleases.** Science. 2009 Jul 24;325(5939):433. doi: 10.1126/science.1172447.

Harrison PT, Hoppe N, Martin U. **Gene editing & stem cells.** Journal of Cystic Fibrosis 2017. http://www.cysticfibrosisjournal.com/article/S1569-1993(17)30975-X/fulltext

Hart SL, Harrison PT. **Genetic therapies for cystic fibrosis lung disease. Current Opinion in Pharmacology.** Curr Opin Pharmacol. 2017 Jun;34:119-124. doi: 10.1016/j.coph.2017.10.006.

Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, Steinfeld I, Lunstad BD, Kaiser RJ, Wilkens AB, Bacchetta R, Tsalenko A, Dellinger D, Bruhn L, Porteus MH. **Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells.** Nat Biotechnol. 2015 Sep;33(9):985-989. doi: 10.1038/nbt.3290.

Hollywood JA, Lee CM, Scallan MF, Harrison PT. **Analysis of gene repair tracts from Cas9/gRNA double-stranded breaks in the human CFTR gene.** Sci Rep. 2016 Aug 25;6:32230. doi: 10.1038/srep32230.

Ingram VM. **Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin.** Nature. 1957 Aug 17;180(4581):326-8.

Jansen R, Embden JD, Gaastra W, Schouls LM. **Identification of genes that are associated with DNA repeats in prokaryotes.** Mol Microbiol. 2002 Mar;43(6):1565-75.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.** Science. 2012 Aug 17;337(6096):816-21. doi: 10.1126/science.1225829.

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. **Identification of the cystic fibrosis gene: genetic analysis.** Science. 1989 Sep 8;245(4922):1073-80.

Kim S, Kim D, Cho SW, Kim J, Kim JS. **Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome research.** 2014;24:1012-9.

Kucherlapati RS, Eves EM, Song KY, Morse BS, Smithies O. **Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA.** Proc Natl Acad Sci U S A. 1984 May;81(10):3153-7.

Lander ES et al. **Initial sequencing and analysis of the human genome.** Nature. 2001 Feb 15;409(6822):860-921.

Lee CM, Flynn R, Hollywood JA, Scallan MF, Harrison PT. **Correction of the ΔF508 Mutation in the Cystic Fibrosis Transmembrane Conductance Regulator Gene by Zinc-Finger Nuclease Homology-Directed Repair.** Biores Open Access. 2012 Jun;1(3):99-108. doi: 10.1089/biores.2012.0218.

Lee CM, Davis TH, Bao G. **Examination of CRISPR/Cas9 design tools and the effect of target site accessibility on Cas9 activity.** Exp Physiol. 2017 Mar 16. doi: 10.1113/EP086043.

Li H, Haurigot V, Doyon Y, Li T, Wong SY, Bhagwat AS, Malani N, Anguela XM, Sharma R, Ivanciu L, Murphy SL, Finn JD, Khazi FR, Zhou S, Paschon DE, Rebar EJ, Bushman FD, Gregory PD, Holmes MC, High KA. **In vivo genome editing restores haemostasis in a mouse model of haemophilia.** Nature. 2011 Jun 26;475(7355):217-21. doi: 10.1038/nature10177.

Lingeman E, Jeans C, Corn JE. **Production of Purified CasRNPs for Efficacious Genome Editing.** Curr Protoc Mol Biol. 2017 Oct 2;120:31.10.1-31.10.19. doi: 10.1002/cpmb.43.

Morales EE, Wingert RA. **Zebrafish as a Model of Kidney Disease.** Results Probl Cell Differ. 2017;60:55-75.

Ormond KE, Mortlock DP, Scholes DT, Bombard Y, Brody LC, Faucett WA, Garrison NA, Hercher L, Isasi R, Middleton A, Musunuru K, Shriner D, Virani A, Young CE. **Human Germline Genome Editing.** Am J Hum Genet. 2017 Aug 3;101(2):167-176. doi: 10.1016/j.ajhg.2017.06.012.

Parry V. **Commit to talks on patient data and public health.** Nature. 2017 Aug 8;548(7666):137. doi: 10.1038/548137a.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. **Genome engineering using the CRISPR-Cas9 system.** Nat Protoc. 2013 Nov;8(11):2281-2308. doi: 10.1038/nprot.2013.143.

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. **Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.** Science. 1989 Sep 8;245(4922):1066-73.

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, et al. **Identification of the cystic fibrosis gene: chromosome walking and jumping.** Science. 1989 Sep 8;245(4922):1059-65.

Sangamo 2017. **Sangamo Announces Treatment of First Patient in Landmark Phase 1/2 Clinical Trial Evaluating In Vivo Genome Editing for MPS II.** https://investor.sangamo.com/press-releases/detail/381/sangamo-announces-treatment-of-first-patient-in-landmark

Sanz DJ, Hollywood JA, Scallan MF, Harrison PT. **Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of CFTR mRNA.** PLoS One. 2017 Sep 1;12(9):e0184009. doi: 10.1371/journal.pone.0184009

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. **DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.** Nature. 2014 Mar 6;507(7490):62-7. doi: 10.1038/nature13011.

Suzuki S, Sargent RG, Illek B, Fischer H, Esmaeili-Shandiz A, Yezzi MJ, Lee A, Yang Y, Kim S, Renz P, Qi Z, Yu J, Muench MO, Beyer AI, Guimarães AO, Ye L, Chang J, Fine EJ, Cradick TJ, Bao G, Rahdar M, Porteus MH, Shuto T, Kai H, Kan YW, Gruenert DC. **TALENs Facilitate Single-step Seamless SDF Correction of F508del CFTR in Airway Epithelial Submucosal Gland Cell-derived CF-iPSCs.** Mol Ther Nucleic Acids. 2016 Jan 5;5:e273. doi: 10.1038/mtna.2015.43.

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. **Highly efficient endogenous human gene correction using designed zinc-finger nucleases.** Nature. 2005 Jun 2;435(7042):646-51.

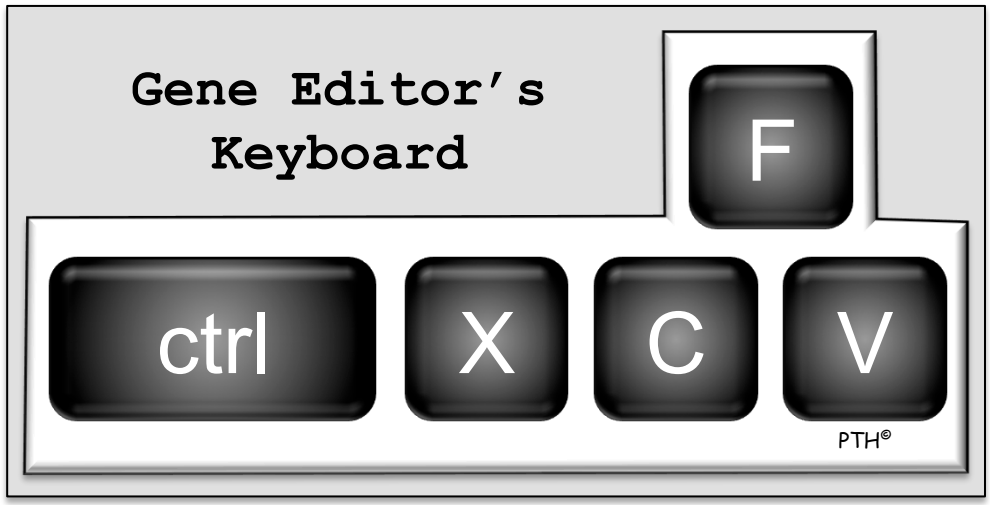
Wang G, Yang L, Grishin D, Rios X, Ye LY, Hu Y, Li K, Zhang D, Church GM, Pu WT. **Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies.** Nat Protoc. 2017 Jan;12(1):88-103. doi: 10.1038/nprot.2016.152.

Wang HX, Li M, Lee CM, Chakraborty S, Kim HW, Bao G, Leong KW. **CRISPR/Cas9-Based Genome Editing for Disease Modeling and Therapy: Challenges and Opportunities for Nonviral Delivery.** Chem Rev. 2017 Aug 9;117(15):9874-9906. doi: 10.1021/acs.chemrev.6b00799.

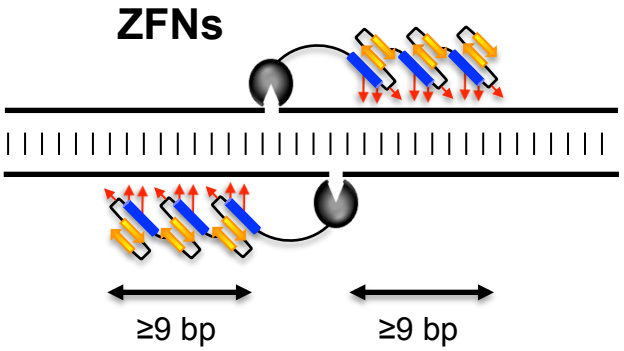
Watson JD, Crick FH. **Genetical implications of the structure of deoxyribonucleic acid.** Nature. 1953 May 30;171(4361):964-7.

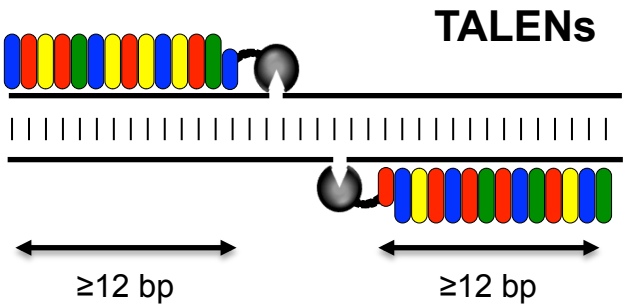
Whitelaw CB, Sheets TP, Lillico SG, Telugu BP. **Engineering large animal models of human disease.** J Pathol. 2016 Jan;238(2):247-56.

Box 1



Box 2


ZFNs
≥9 bp
≥9 bp


TALENs
≥12 bp
≥12 bp

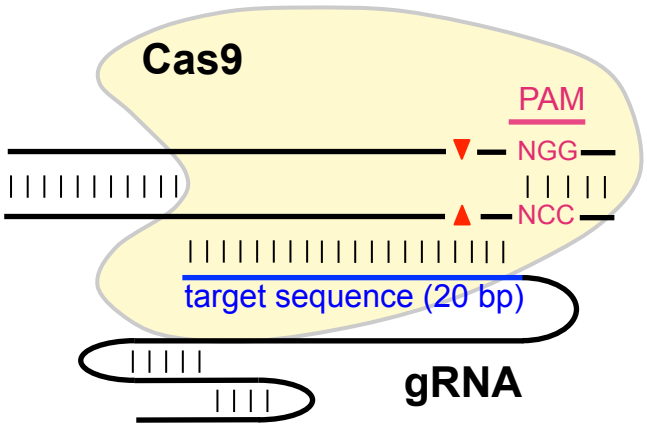

Cas9
PAM
NGG
NCC
target sequence (20 bp)
gRNA

Figure 1

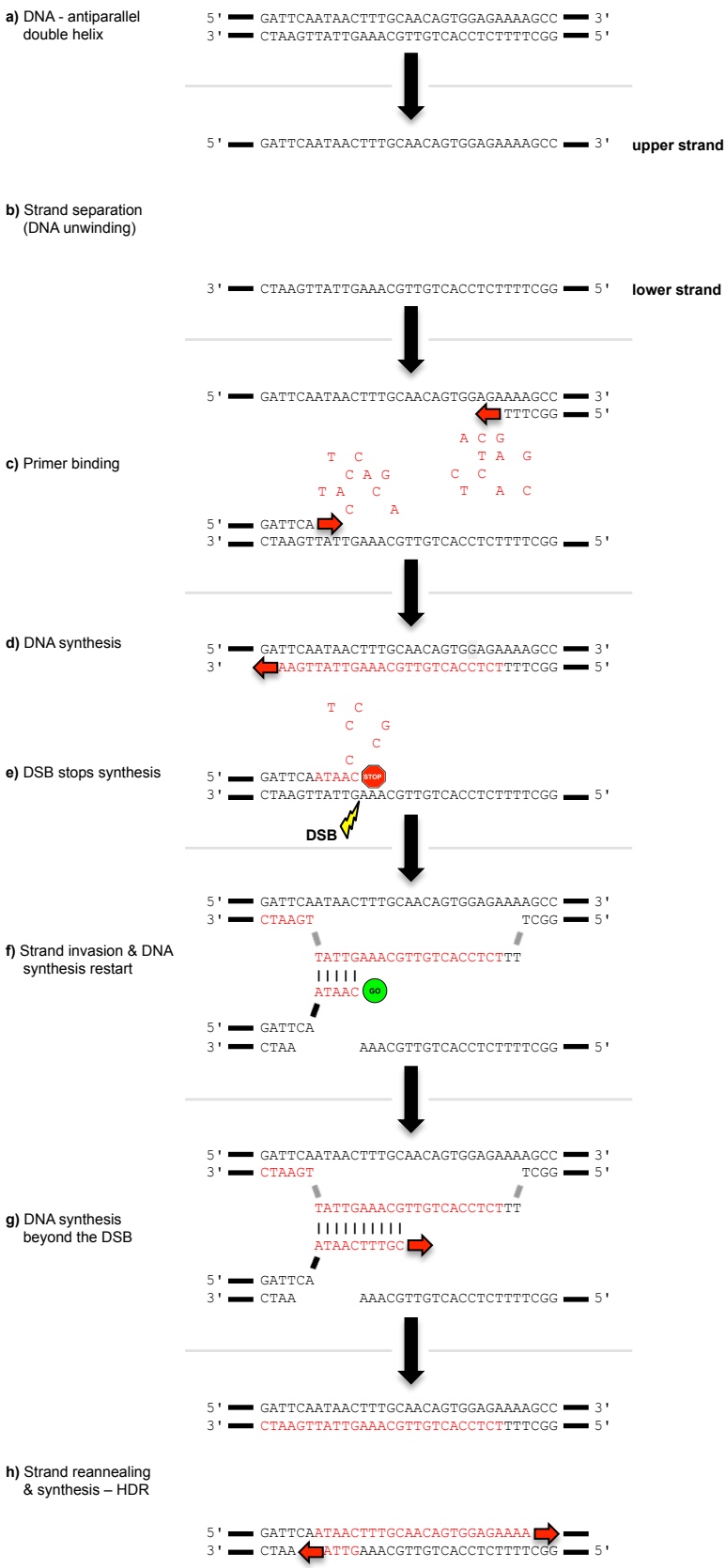

a) DNA - antiparallel double helix
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
3' CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
upper strand
b) Strand separation (DNA unwinding)
3' CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
lower strand
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
TTTCGG 5'
c) Primer binding
5' GATTCA
3' CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
d) DNA synthesis
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
3' AAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
e) DSB stops synthesis
5' GATTCAATAAC STOP
3' CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
DSB
f) Strand invasion & DNA synthesis restart
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
3' CTAAGT TCGG 5'
TATTGAAACGTTGTCACCTCTTT
ATAAC GO
5' GATTCA
3' CTAA AAACGTTGTCACCTCTTTTCGG 5'
g) DNA synthesis beyond the DSB
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
3' CTAAGT TCGG 5'
TATTGAAACGTTGTCACCTCTTT
ATAACTTTGC
5' GATTCA
3' CTAA AAACGTTGTCACCTCTTTTCGG 5'
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
3' CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
h) Strand reannealing & synthesis – HDR
5' GATTCAATAACTTTGCAACAGTGGAGAAAA
3' CTAA ATTGAAACGTTGTCACCTCTTTTCGG 5'

Figure 2

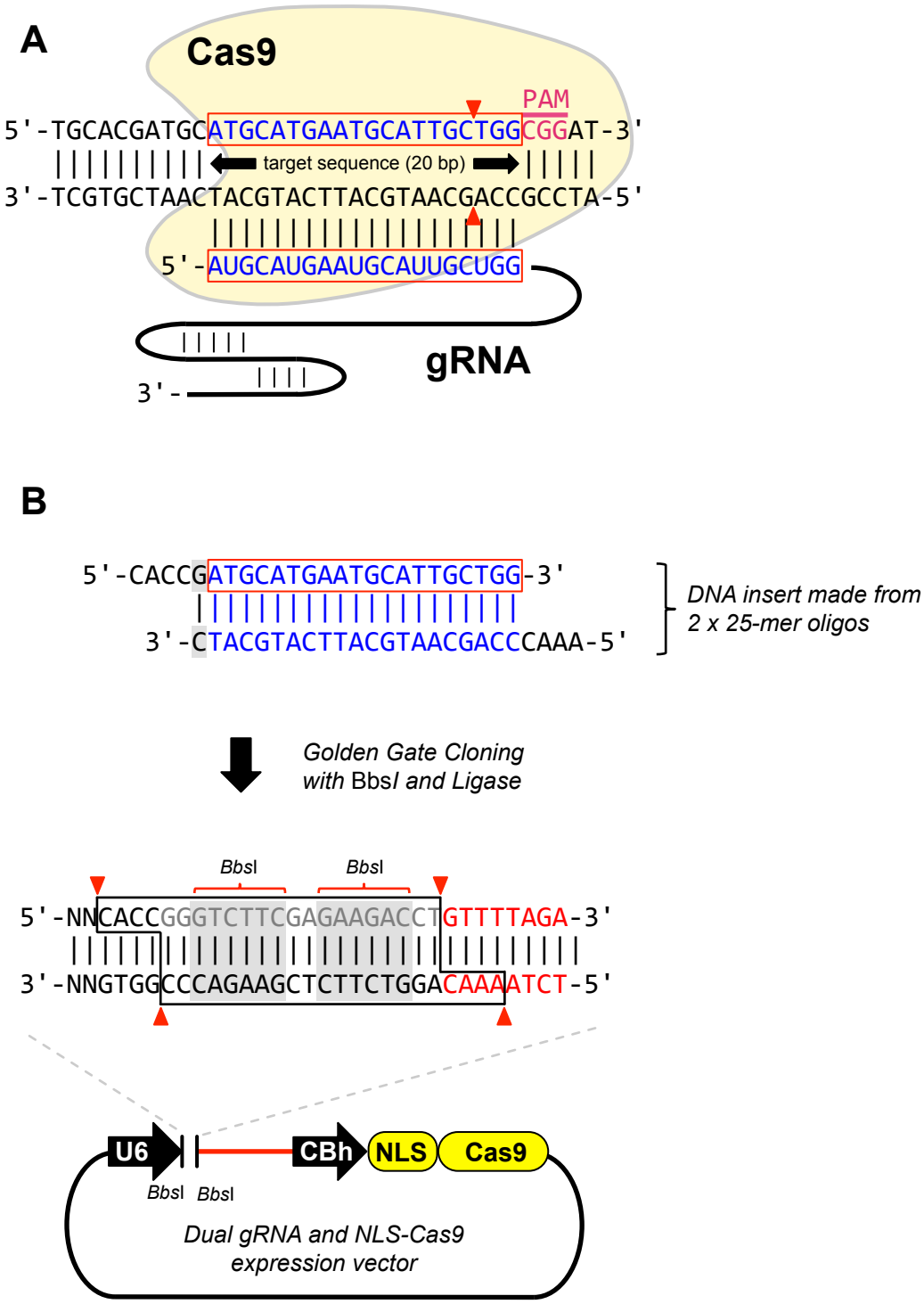

A
Cas9
PAM
5'-TGCACGATGCATGCATGAATGCATTGCTGGCGGAT-3'
target sequence (20 bp)
3'-TCGTGCTAACTACGTACTTACGTAACGACCGCCTA-5'
5'-AUGCAUGAAUGCAUUGCUGG
gRNA
3'-
B
5'-CACCGATGCATGAATGCATTGCTGG-3'
3'-CTACGTACTTACGTAACGACCCAAA-5'
DNA insert made from
2 x 25-mer oligos
Golden Gate Cloning
with BbsI and Ligase
BbsI
BbsI
5'-NNCACCGGGTCTTCGAGAAGACCTGTTTTAGA-3'
3'-NNGTGGCCCAGAAGCTCTTCTGGACAAAATCT-5'
U6
CBh
NLS
Cas9
BbsI
BbsI
Dual gRNA and NLS-Cas9
expression vector

Figure 3

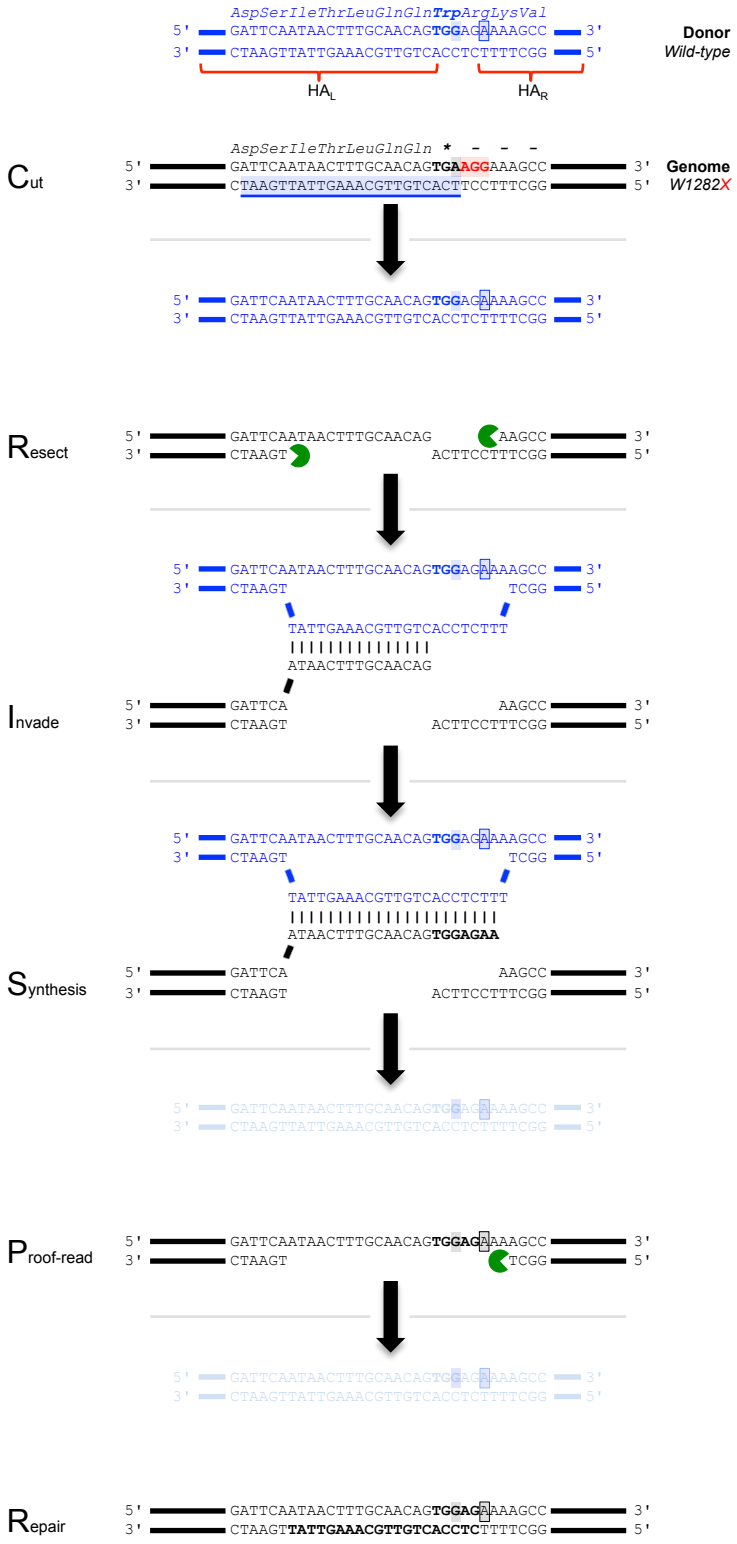
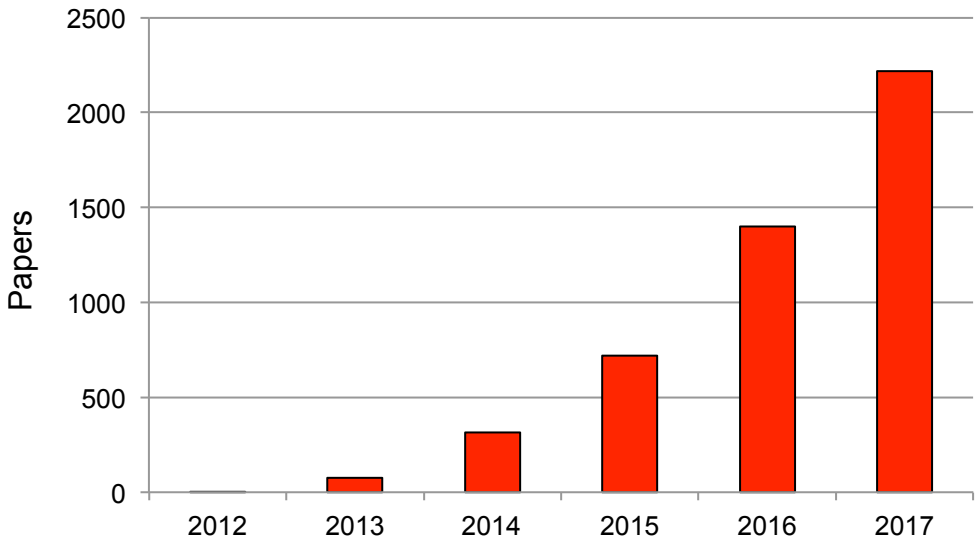

AspSerIleThrLeuGlnGlnTrpArgLysVal
5' GATTCAATAACTTTGCAACAGTGGAGAAAAGCC 3'
3' CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG 5'
Donor
Wild-type
HA_L
HA_R
Cut
AspSerIleThrLeuGlnGln * - - -
5' GATTCAATAACTTTGCAACAGTGAAGGAAAGCC 3'
3' CTAAGTTATTGAAACGTTGTCACTTCCTTTCGG 5'
Genome
W1282X
Resect
Invade
TATTGAAACGTTGTCACCTCTTT
ATAACTTTGCAACAG
Synthesis
ATAACTTTGCAACAGTGGAGAA
Proof-read
Repair
CTAAGTTATTGAAACGTTGTCACCTCTTTTCGG

Figure 4


Papers
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017