CRISPR systems: what's new, where next?

Ashley Parkes, **Fiona Kemm**, **Liu He** and **Tom Killelea** (University of Nottingham, UK)

The genetic signature of natural CRISPR-Cas systems were first noted in a 1989 publication and were characterized in detail from 2002 to 2007, culminating in the first report of a prokaryotic adaptive immune system. Since then, CRISPR-Cas enzymes have been adapted into molecular biology tools that have transformed genetic engineering across domains of life. In this feature article, we describe origins, uses and futures of CRISPR-Cas enzymes in genetic engineering: we highlight advances made in the past 10 years. Central to these advances is appreciation of interplay between CRISPR engineering and DNA repair. We highlight how this relationship has been manipulated to create further advances in the development of gene editing.

CRISPR - where we are now

The year 2022 will mark the 10-year anniversary of the initial sparks that ignited the 'CRISPR explosion'. CRISPR – clustered regularly interspersed short palindromic repeats – had caught the attention of researchers prior to 2012, initially proposed to be involved in prokaryotic chromosome partition, or DNA repair, then identified as a system of prokaryotic adaptive immunity: however, it was the landmark papers and development of CRISPR-Cas9-based genetic editing through work led by the Charpentier, Church, Doudna, Šikšnys and Zhang labs that gained widespread recognition as a revolutionary new molecular biology tool. The breakthroughs show no sign of abating with CRISPR surpassing 5000 mentions in individual pieces of literature in 2019 (Figure 1) and have resulted in award of the Nobel Prize for chemistry to Jennifer Doudna and Emmanuelle Charpentier in 2020.

In the years since, CRISPR editing has been adapted (e.g., EvolvR, prime editing) and refined (e.g., Cas9-CRISPRi and -CRISPRa) into exciting applications. These have impacted in revealing molecular mechanisms in disease modelling – notably the cancer dependency map (DepMap) for discovering synthetic lethality between DNA damage repair and signalling systems critical for identifying novel cancer treatment targets. Animal trials using CRISPR editing for genetic therapy *in vivo* have revealed promising data for progress in treating muscular dystrophy. The system has also found uses within agriculture as a tool for crop disease resistance and improved yields. DETECTR and SHERLOCK systems, Cas12- and Cas13-based tools (Cas9 alternatives), respectively, have been deployed for rapid identification of Zika and SARS-COV-2 infections. CRISPR 'gene drives' for the eradication of disease by targeting vectors such as mosquitoes that transmit malaria, an application with strong ethical considerations.

Molecular biology of CRISPR systems

Naturally occurring CRISPR are specialized regions of prokaryotic chromosomes that comprise a 'leader' (promoter), genes encoding Cas (CRISPR-associated) proteins and CRISPR DNA, constituting repeat DNA sequences juxtaposed with 'spacer' regions. Detectable in 45% of known bacterial clades and 87% of archaea, diversity of CRISPR systems is simplified into two classes: class I that degrades MGEs via multiple protein-RNA effectors (e.g., CASCADE) and class II systems using a single protein (e.g., Cas9, Cas12). In some respects, CRISPR systems are analogous to the well-characterized restriction modification system. Both are highly diverse prokaryotic immune defences that target invading phage DNA, though restriction enzymes are promiscuous and less specific: if restriction modification could be considered a prokaryote's system of innate immunity, then CRISPR could be considered its system of adaptive immunity.

The major steps of CRISPR immunity within prokaryotes have been covered numerous times in many review articles – the host cell Cas proteins 'capture' and 'integrate' MGEs into a CRISPR locus as a new 'spacer' ('adaptation', Figure 2) generating an extended CRISPR locus that is transcribed to RNA (crRNA) that after some trimming targets any repeat visit from the MGE for destruction by 'interference' (Figure 3).

Adaptation, in context of CRISPR systems, generates a genetic memory of encountered MGEs. The most well-known enzyme of interference leading to MGE degradation and providing new DNA for adaptation is Cas9; interference reactions by Cas9 form the basis of commonly used CRISPR-Cas9 genetic editing reactions. This crRNA-guided interference of DNA by Cas9 caught the attention of researchers due to its ability to generate double-strand breaks within DNA: in the context of

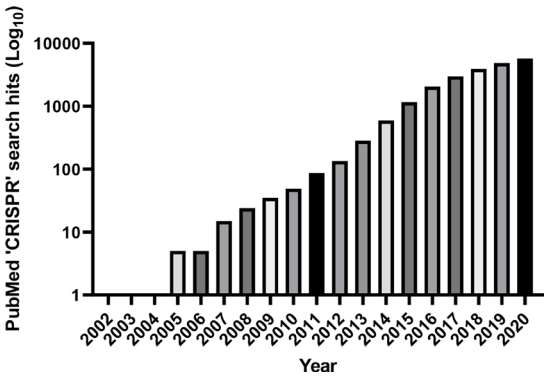

Figure 1. PubMed search results filtered by abstract for years 2002 until 2020, using the term 'CRISPR' as a keyword. CRISPR footprint within literature has increased from one reference in the 2002 paper of Jansen et al that coined the term 'CRISPR' to over 5000.

CRISPR immunity these lead to destruction of MGEs. In the context of genome editing, they provide a way to manipulate chromosomes for gene editing.

Until 2012, gene editing had focused on meganucleases, and TALEN or zinc finger nucleases. Now somewhat outdated, these methods were some of the first to combine programmable specific DNA binding with non-specific DNA cleavage. Both systems required the laborious custom design and creation of a new protein for each DNA sequence to be targeted. CRISPR was thus attractive as, by simply changing the crRNA sequence, one could target any DNA region containing a corresponding PAM, with estimates that the classical *Streptococcus pyogenes* Cas9 motif (NGG) appears 1.6×10^8 times across the human genome. In a landmark study, the Doudna lab at UC Berkeley demonstrated functionality of a dual tracrRNA:crRNA system in inducing double-strand breaks within DNA. The Zhang lab subsequently adopted and showed functionality of this design *in vitro* within mammalian cells. This has enabled genetic editing across prokaryotic and eukaryotic cells, with five licensed UK medical trials (phase one) currently active using CRISPR-Cas9-modified cells as therapy.

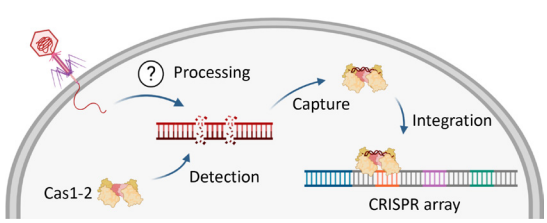

Figure 2. Overview of naive adaptation by CRISPR-Cas systems. MGE, here in the form of a bacteriophage, injects genetic material (red) into host cells. MGE DNA is identified and processed by as-yet unidentified mechanisms and captured by the Cas1-Cas2 complex, prior to integration into the CRISPR array.

DNA repair – the processes that make CRISPR editing work

The application of Cas9-gRNA complexes for genetic editing is the trigger for processes completed by endogenous DNA repair systems. DNA repair does the heavy lifting but also adds considerable unpredictability and complexities to genome editing reactions, particularly those editing reactions that rely on homology-directed repair (HDR) for insertions and non-homologous end-joining (NHEJ) for deletions (see Figure 4 for a simplified overview).

HDR uses a sequence of homologous DNA to repair the double-strand break in cell cycle stages when a template strand is available. The mechanisms and extensive network of proteins involved in HDR have been exploited by researchers to deliver the more classic idealized form of CRISPR editing, termed 'knock-in gene editing', whereby a DNA template possessing homology arms designed around the site of the double-strand break is included in the CRISPR reaction to insert and modify DNA as desired.

NHEJ is distinct from HDR by not requiring significant complementarity with a homologous DNA sequence for repair; consequently, NHEJ can occur at any stage of the cell cycle and has historically been considered to be more error-prone than HDR, often resulting in insertion/deletion events at the double-strand break. However, more recent studies have begun to demonstrate genomic instabilities promoted by HDR, notably associated within the pathway of break-induced replication (BIR). Nevertheless, NHEJs ability to generate indels has been exploited to create gene knockouts via CRISPR and opened genetic editing to a wider variety of labs.

DNA repair and CRISPR, overlapping processes in genome dynamics

Model organisms have served as high-quality models for understanding human molecular pathways. But contradictions have arisen; in yeast, for instance, Rad52 (a HDR annealase) operates within double-strand break repair whilst in humans Rad52 plays a subtler role in multiple guises, to the extent that it was previously considered to play a back-up role. Likewise, the BRCA2 family of proteins, pivotal in human recombinational DNA repair, is completely absent from budding yeast. Models in *Drosophila* have also shown inconsistencies between orthologues, like the transcriptional regulator Yki and its human counterpart Yap, a known oncogene. Thus for validating functional roles of suspected human

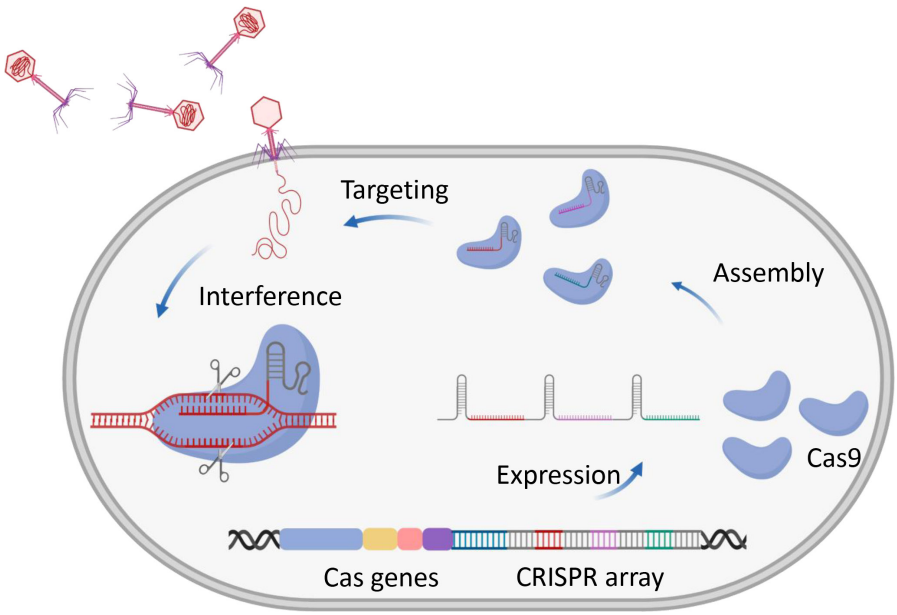

Figure 3. Interference by CRISPR-Cas9. Cas9 assembles as a ribonucleoprotein complex with processed crRNA and tracrRNA. The Cas9-RNA complex targets invading MGE DNA at a region complimentary to the crRNA, forming a stable R-loop, with subsequent cleavage of both strands of the MGE DNA. DNA damage initiates intrinsic repair mechanisms in host cells, leading to degradation of invader DNA.

orthologues, *in vitro* studies with human cells are necessary: studies made more accessible with the advent of CRISPR.

Understanding interplay between repair pathways and Cas9 has informed the development of CRISPR-based gene-editing techniques. Whilst knockout gene editing through NHEJ is very accessible, attempts to use HDR to achieve more precise gene editing show low efficacies. Synchronizing and stalling the cell cycle to encourage repair machinery towards HDR has been successful; however, even under these favourable conditions the machinery of HDR acts in direct competition with NHEJ. Researchers have negotiated these issues and improved efficiencies further by suppressing proteins crucial to NHEJ and creating fusions of the Cas9 effector with the HDR protein CtIP, localizing CtIP to the site of the break, biasing repair towards HDR. The Cas9 protein itself has been the subject of mutagenesis studies to favour HDR, such as in Cas9 nickases: a mutated form of Cas9s nuclease domain such that DNA is only nicked rather than fully cut. By coupling a gRNA with a second gRNA further downstream, one

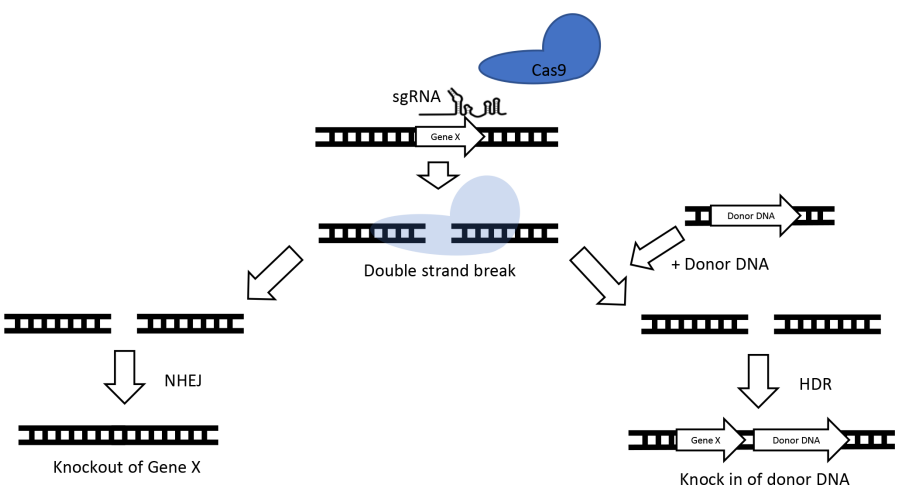

Figure 4. A simplified summary CRISPR-Cas9-based gene editing. Cas9-gRNA creates a DNA double-strand break that recruits proteins of non-homologous end-joining and homology-directed repair to create a deletion or insert a piece of donor DNA, respectively.

could create a double-strand break containing overhangs ideal for implementing a cassette through HDR whilst also reducing off-target effects.

Whilst promising, none of these methods represent a panacea for improving HDR and many would be difficult, if not impossible, to carry out *in vivo*. Fusing Cas9 to CtIP, e.g., would compromise the capacity available for additional cofactors needed for editing within the limited size of delivery vectors. However, there are still substantial gaps in our knowledge of human DNA repair that may yet inform the design of editing systems. The Fanconi anaemia repair pathway is one such pathway: implicated in Cas9-based genetic editing but with the understanding of the interactions still limited. CRISPR-based tools can be readily applied to help understand these poorly understood pathways, informing design of CRISPR systems in addition to other fields.

Next for CRISPR in genetic editing?

After 10 years of developments CRISPR hasn't quite yet matured into a precision genetic surgery kit for widespread medical use. Complex interactions between CRISPR-based editing processes and human DNA repair systems render efficiencies low and unpredictable; and destabilizing factors urge caution for practical use in treating disease. For example, the induction of off-target effects in unrelated genes remains a problem for CRISPR-based gene editing. Then, even when the technical aspects have been deemed sufficiently validated for medicinal use, ethical considerations mean the use of CRISPR-based gene editing in industry will remain a hot topic for years to come.

Two developments are promising now: base editing and prime editing. Base editing systems use a modified Cas9, notably catalytically dead Cas9 (dCas9) and nCas9, fused with DNA base-modifying enzymes to precisely edit a single base: these include cytidine base editors (CBEs) that create C>T conversions and adenine base editors (ABEs) that permit A>G conversions. Whilst showing great potential for correcting single nucleotide polymorphisms, these were limited by only allowing transition mutations (purine to purine/pyrimidine to pyrimidine) and the necessity for a PAM sequence positioned 13–17 nucleotides away from the target site. Prime editing, which uses an extended gRNA (pegRNA)-guided reverse transcriptase, represented a further breakthrough as it removed necessity for this PAM positioning, whilst demonstrating capabilities for indel mutations and all 12 base conversions. Prime editing also demonstrated reductions in off-target effects and has been optimized to similar efficiencies as base editing systems; recent examples have seen improved efficiencies via fusions with the DNA binding domain of the recombinational DNA repair protein Rad51.

The optimization of CRISPR-Cas9 gene-editing tools over the last decade has created a rush for effective uses in both biotechnology and diagnostics. However, the following question remains: will efforts to optimize the system for therapeutic applications be successful? Or could a new gene-editing technology arrive and relegate CRISPR-based genetic editing to the annals of history, like the zinc finger nucleases and TALENs that came before it? Both remain open questions, but one thing is clear: the future of CRISPR-based gene editing is intertwined with DNA repair.

Acknowledgements

The authors are all Early Career Researchers working on CRISPR molecular biology in Ed Bolt's group, supported by grant funding from The BBSRC, and Industrial sponsors.

Funding

All of the authors are Early Career Researchers in Ed Bolt's research group at The University of Nottingham. Work involving CRISPR is supported by BBSRC grant BB/T006625/1, BBSRC DTP studentships, and industrial-funded discovery grants. ■

Further reading

- Charpentier, E., Elsholz, A. and Marchfelder, A. (2019) CRISPR-Cas: more than ten years and still full of mysteries. *RNA Biol*. **16**, 377–379. DOI: 10.1080/15476286.2019.1591659
- Cubbon, A., Ivancic-Bace, I. and Bolt, E.L. (2018) CRISPR-Cas immunity, DNA repair and genome stability. *Biosci. Rep*. **38**, BSR20180457. DOI: 10.1042/BSR20180457
- Doudna, J.A. (2020) The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236. DOI: 10.1038/s41586-020-1978-5
- Jinek, M., Chylinski, K., Fonfara, I. et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821. DOI: 10.1126/science.1225829

(Continued)

Further reading (*Continued*)

- Makarova, K.S., Wolf, Y.I., Iranzo, J. et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83. DOI: 10.1038/s41579-019-0299-x
- Ran, F.A., Hsu, P.D., Wright, J. et al. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308. DOI: 10.1038/nprot.2013.143
- Richardson, C.D., Kazane, K.R., Feng, S.J. et al. (2018) CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* 50, 1132–1142. DOI: 10.1038/s41588-018-0174-0

Ashley Parkes is a PhD student in a 4-year BBSRC Doctoral Training Partnership (PhD). Ashley is investigating the function of DExD helicases through biochemistry techniques and CRISPR-based gene knockouts. Email: stxap25@nottingham.ac.uk

Fiona Kemm is an MRes student interested in the molecular mechanisms behind CRISPR-Cas systems, particularly the functionalities of proteins involved in adaptation. Email: mbyfk3@nottingham.ac.uk

Liu He is a vice chancellor's scholarship-funded PhD student. Her research focuses on molecular biology of Cas3, understanding its structural compactness and functional variability. Email: Liu.He1@nottingham.ac.uk

Tom Killelea is a post-doctoral research fellow on the project CRISPR Adaptation: How Is DNA Captured? Tom investigates mechanism and interaction of bacterial adaptation complexes with other cellular processes, including DNA repair. Email: mbztk@nottingham.ac.uk

