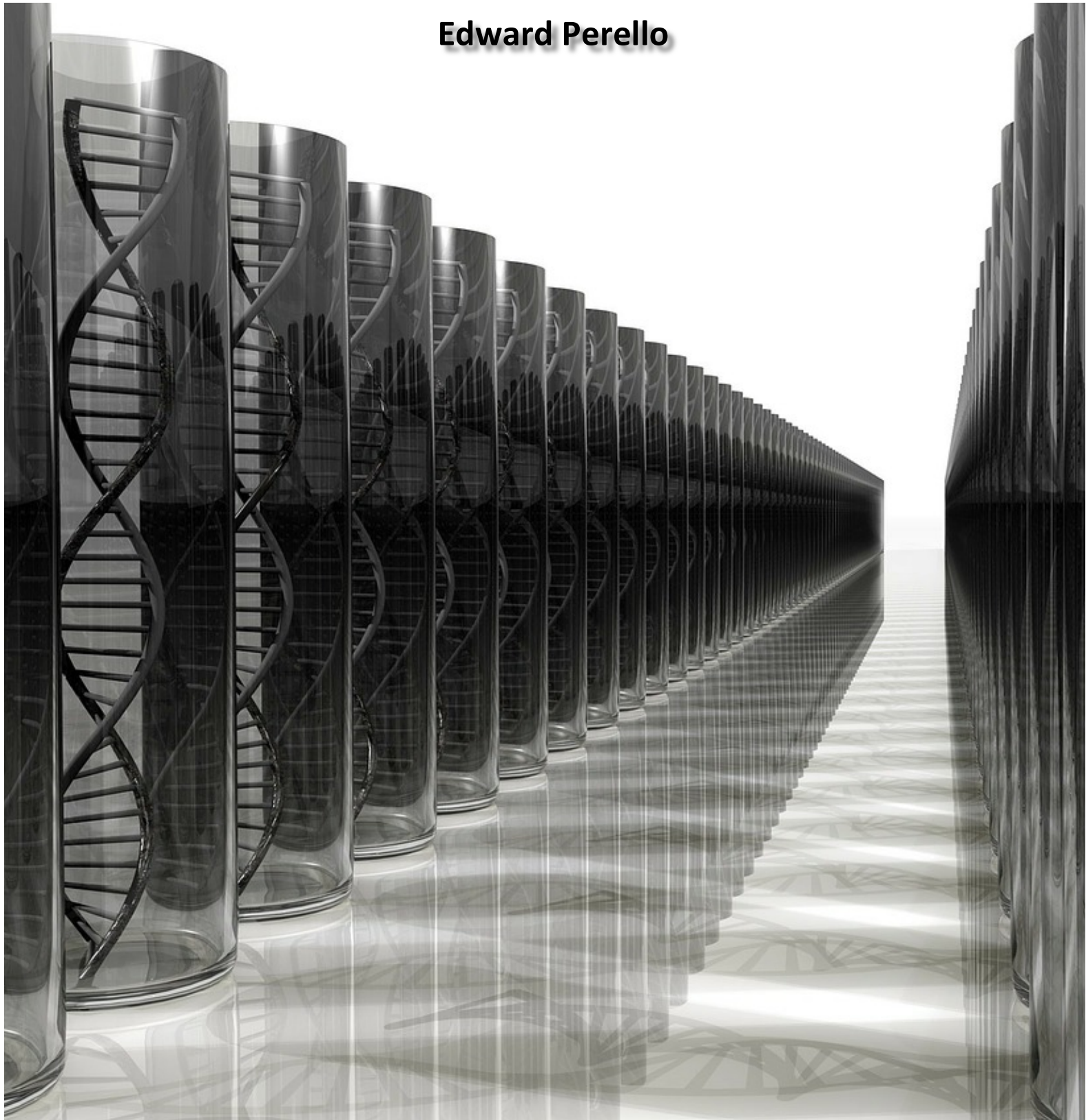


# CRISPR GENOME EDITING: A TECHNICAL & POLICY PRIMER

Edward Perello



The overarching goal of this study is to present policy options and recommendations to key stakeholders. In the design of these options and recommendations, the research team focused on how to manage the often-competing demands of promoting innovation and preventing misuse, and how to adapt current, or create new, governance mechanisms to achieve these objectives.

This paper seeks to provide clarity and context to the *field*, *process*, and *tools* of contemporary genome editing to enable biologists and non-biologists to contribute to policy development and oversight by individuals, organizations, agencies, and governments alike. The paper explores genome editing tools, the processes in which they are used, their capabilities and limitations, the social context influencing their development and use, and considerations for their effective control. Close attention is paid to the CRISPR tool, but the goal has been to sufficiently generalize our considerations to ensure relevance to past and future tools, and inform ongoing research in this area.

The four study leads and three research assistants for *Editing Biosecurity* were assisted by a core research group of fourteen subject-matter experts with backgrounds in the life sciences, industry, policy, ethics, and security. The centerpiece of the study was three invitation-only workshops that brought together the core research group for structured discussions of the benefits, risks, and governance options for genome editing. To support these workshops, the study leads prepared two working papers on risk assessment and governance and commissioned five issue briefs on key topics.

All of these working papers and issue briefs are available at the project's website:  
<https://editingbiosecurity.org/>.

A list of project participants can be found in the project's final report, *Editing Biosecurity: Needs and Strategies for Governing Genome Editing*, **which is available at:**  
[www.editingbiosecurity.org](http://www.editingbiosecurity.org).

# EDWARD PERELLO BIOGRAPHY

Edward Perello is the Principal Researcher for Arkurity, a boutique consulting firm conducting research on public policy challenges in synthetic biology, conservation biotech, and biosecurity. Edward's research interests include the oversight of human genome editing, state and non-state actor development of biological capabilities, and the application of synthetic biology to ecological challenges. He is a Research Fellow at George Mason University, where he works on security policy for genome editing tools. He currently serves on the IUCN Task Force on Synthetic Biology and Biodiversity Conservation, and is working with conservation groups to realize new opportunities for biotechnology in ecosystem restoration. He previously founded Desktop Genetics, a CRISPR biotechnology company, and served as Chief Business Officer for six years. Edward co-chaired the iGEM software committee for two years and is an alumnus of the ELBI biosecurity and SynBio LEAP fellowships.

## **Affiliations**

George Mason University, Institute for Philosophy and Public Policy  
Arkurity Ltd

## **Contact email**

[edward.perello@arkurity.com](mailto:edward.perello@arkurity.com)

## **Acknowledgements**

The author wishes to thank Joanna Kamens at Addgene, Greg Koblentz, Jesse Kirkpatrick, and Sarah Denton at George Mason University, Megan Palmer at Stanford University, Reza Rohani at University of Oxford, and Rudiger Trojok at Technical University Munich.

## TABLE OF CONTENTS

Glossary of Terms .....	1
Section 1: Genome Editing at a Glance .....	4
Section 2: Drivers of CRISPR Adoption .....	18
Section 3: CRISPR - Embodiment & Processes .....	27
Section 4: Governance for CRISPR .....	44
Section 5: What's Next? .....	62
References .....	64

# Glossary of Terms

**Bavarian Health and Food Safety Authority (LGL)** - the expert body of Bavaria/Germany for food safety, health, animal health, occupational health and safety/product safety.

**Binding domain** - a protein element that binds to a particular molecule.

**Bioinformatics** – an interdisciplinary field that deploys computational approaches to understand biology.

**Biosafety level (BSL)** - a set of biocontainment precautions required to isolate dangerous biological agents in an enclosed laboratory facility. The levels of containment range from the lowest biosafety level 1 (BSL-1) to the highest at level 4 (BSL-4).

**CRISPR** - a microbial immune system that has been repurposed as a genome editing tool.

**Delivery** - the process by which a molecule is introduced to a particular cell or organism.

**Delivery vector** - the physical format through which a molecule is delivered to a cell or organism.

**dsDNA** - double stranded DNA.

**Foreign DNA** - exogenous DNA deliberately introduced to an organism for the purpose of genetic engineering.

**Genome** - the complete set of genetic information found in an organism's chromosomes.

**Genomic DNA** - most organisms have the same genomic DNA in each of their cells, with a complete genome said to be unique to an individual other than in the case of clones (identical twins or cuttings). Genomic DNA may refer to an individual's unique genome, or a reference genome for a species. Genomic DNA is generally the target of genome editing technologies.

**Genome Editing Vector (GEV)** - a delivery vector used to deliver essential components of a genome editing tool into a cell.

**Genetic transformation** - the process by which the genetic makeup of a cell is altered by taking up DNA from the environment. May simply be referred to as transformation.

**Genetically modified organism (GMO)** - any organism whose genetic material has been altered using genetic engineering techniques.

**Genome editing reagents** - a general term used herein to describe the wetware materials required for a genome editing procedure.

**Genotype** - the set of genes responsible for a particular trait or set of traits.

**Guide / guide RNA** - a nucleic acid based binding domain used in CRISPR genome editing.

**High-throughput screening (HTS)** - a method employing automation to conduct highly parallelized experimentation, especially in drug discovery and functional genomics.

**Homology** - describes the similarity of nucleic acid sequences, with high homology indicating highly similar sequences for DNA, RNA or protein.

**Homology-directed repair (HDR)** - a mechanism in cells that allows double-stranded breaks in DNA to be repaired using a template with homology to the cut site.

**International Gene Synthesis Consortium (IGSC)** - an industry-led group of gene synthesis companies and organizations formed to design and apply a common protocol to screen both the sequences of synthetic gene orders and the customers who place them.

**Mosaic** - a population of cells or an organism with a variable genotype.

**Next generation sequencing (NGS)** - also known as high-throughput sequencing, is a catch-all term used to describe a number of different modern sequencing technologies.

**Non-homologous end joining (NHEJ)** - a mechanism in cells that allows double-stranded breaks in DNA to be repaired without the availability of a homologous template.

**Nuclease** - an enzyme known to cleave nucleic acid sequences.

**Oligonucleotide** - a short string of nucleic acids, sometimes referred to as an oligo.

**Payload** - the entire set of materials or reagents or GEVs and additional components delivered to a cell during a genome editing procedure.

**Phenotype** - the physical expression or characteristics derived from a genotype.

**Plasmid** - circular non-genomic DNA vectors that may be used to transfer extrachromosomal DNA between organisms, or otherwise used as a GEV.

**PAM** - stands for Protospacer Adjacent Motif, a recognition site upon which an RGEN may bind.

**Reference genome** - a representative approximation of an organism's DNA, used to understand an average genome for a particular species.

**Restriction fragment length polymorphism (RFLP)** - a technique that exploits variations in homologous DNA sequences to obtain a visual readout using PCR, restriction enzymes and gel electrophoresis.

**RNA-guided endonuclease (RGEN)** - a nuclease that is guided to its target by way of a programmable RNA element, for instance a guide RNA.

**Transcription activator like effector nuclease (TALEN)** - restriction enzymes that can be engineered to cut specific sequences of DNA in the genome.

**Transgenic** - denoting an organism that contains genetic material into which DNA from an unrelated organism has been artificially introduced.

**Wetware** – a term in the vein of hardware and software, referring to the tangible and intangible biological and molecular elements or information contained in a cell or biological system. May denote living cells, or elements of them used as reagents in experiments.

**Zinc finger nuclease (ZFN)** - artificial restriction enzymes that can be engineered to cut specific sequences of genomic DNA.

# SECTION 1: GENOME EDITING AT A GLANCE

## Framing Genome Editing

Precise and easy genetic modification has long been an objective of the molecular biology community. The dawn of genetic engineering was heralded by the development of the first engineered bacteria by means of introducing *foreign DNA* from one organism into another (Cohen et al. 1973). Early techniques built on these *transgenic* approaches, but remained frustratingly slow, with technical limitations that prevented researchers from executing anything but simple modifications.

In the 45 years since the first transgenic organisms were developed using rudimentary techniques, the field of genetic engineering has changed, the tools have evolved and our capabilities have advanced. In recent years, *genome editing* has emerged as a lively discipline of genetic engineering, making use of successive generations of simple and flexible tools that allow a modern molecular biologist to perform an almost unlimited range of alterations to the genomic DNA in an organism's cells. Genomic DNA forms the bulk of the genetic sequence information of an organism, and may be heritable. Genome editing is a biology capability concerned with the permanent modification and manipulation of this genomic DNA and differs from conventional genetic engineering techniques which typically manipulate or insert non-genomic DNA in a temporary, non-heritable manner.

To appreciate its complexity, it is helpful to approach genome editing from a few perspectives:

- The genome editing **field** comprises the entire set of activities, technologies, cultural norms, and economics associated with the use of these techniques.
- Genome editing **processes** are the generalizable sets of technical operations, not limited to the use of genome editing tools themselves, but other technologies and procedures that are essential in planning, executing, and measuring the outcome of a genome editing procedure.
- Genome editing **tools** are the specific molecular methods that are used to alter an organism's DNA. They may be used in conjunction with other tools, and as part of larger processes. The most well-known of these tools is called CRISPR.

Over the years, the field has produced increasingly useful tools, and has shifted as more capabilities spawn new *processes* involving genome editing. Developments in all three areas of genome editing are accelerating in pace, advancing biologists' technical capabilities, and lowering the costs of game-changing achievement in the life sciences.

This paper seeks to provide clarity and context to the *field*, *process*, and *tools* of contemporary genome editing to enable biologists and non-biologists to contribute to policy development and oversight by individuals, organizations, agencies, and governments alike. The paper explores genome editing tools, the processes in which they are used, their capabilities and limitations, the social context influencing their development and use, and considerations for their effective control. Close attention is paid to the CRISPR tool, but the goal has been to sufficiently



generalize our considerations to ensure relevance to past and future tools, and inform ongoing research in this area.

## Genome Editing at a Glance

Genome editing generally involves the genomic DNA of a living cell being cut and subsequently modified using a *nuclease* protein that is well-suited to this task. The nuclease protein is typically guided by a programmable element that acts as a targeting molecule, or customized binding domain.

A key concept in genome editing is programmability of this custom binding domain, which results from computationally predicting the molecular interactions that might occur between it and a target, and selecting the best sequence of nucleotides to guide the nuclease to that target. Much is possible when DNA, RNA, and protein are brought into close proximity, and genome editing tools can leverage knowledge of biology for impressive effect by exploiting these interactions in complex ways.

A typical genome editing approach sees short DNA or RNA strands called oligonucleotides designed using bioinformatic techniques to determine how to precisely edit “target” sequences in a gene or other important genetic element. Oligonucleotides of a length  $n$ -mer (e.g. a 20-mer having a length of 20 nucleotides) are chemically synthesized as a string and then delivered into cells in culture through means of a genome editing vector (GEV). The GEV contains a payload of protein, DNA, or RNA, which somehow instructs the cell to modify its own genomic DNA. This is typically done by having the genome editing machinery cut the cell’s genome at the target site. When cells detect these changes in their genomes, further editing events occurs by way of DNA repair processes that are either stimulated automatically, or by means of some other payload elements that are co-delivered alongside, or within, the GEV.

Millions of cells can be involved a typical genome editing procedure, and GEVs can be introduced to any of these cells in a population. As such it is helpful to keep in mind that these events are much like a stochastically-driven chemical reaction, with a rate of reaction, a rate of successful editing, and also a rate of unsuccessful editing. Specific target DNA sequences may occur many times throughout a cell’s genomic DNA, complicating consistent or specific editing. This creates a risk that other mission-critical elements of a genome might be accidentally modified causing off-target effects with unknown implications. On the other hand, most cells will fail to be edited at all.

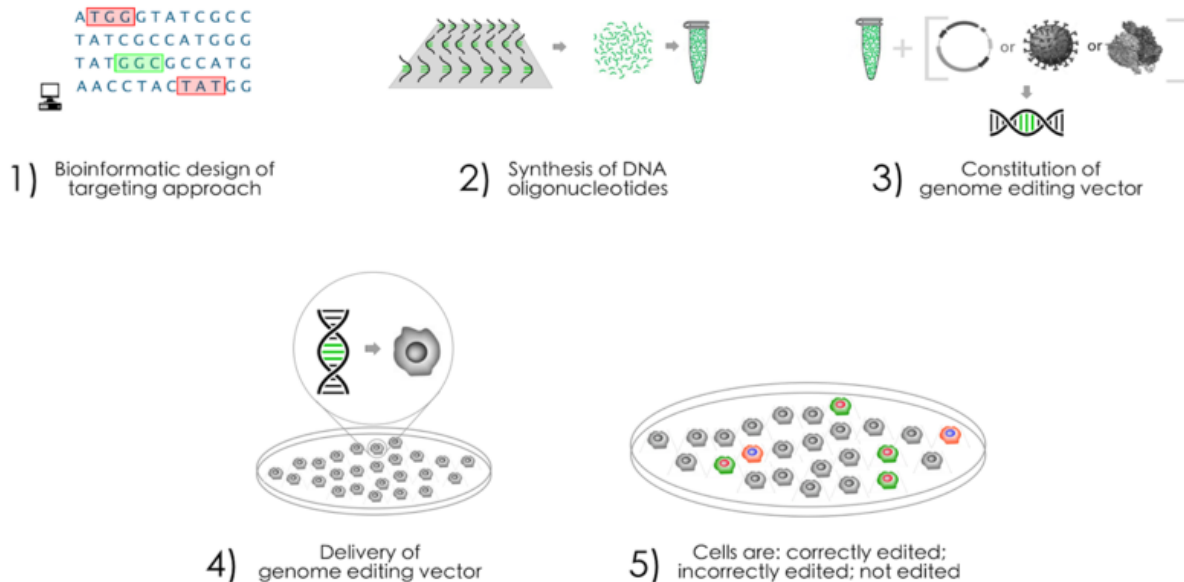


Figure A. Schematic showing basic events occurring in an ideal genome editing process: 1) bioinformatic design with selection of targets and deliberate avoidance of off targets; 2) synthesis and manufacturing of novel programmable oligonucleotides that correspond to bioinformatic design; 3) constitution of a genome editing vector, which pairs the programmable oligo with a DNA plasmid, virus or protein, each one carrying a specific set of instructions for the cell to edit its own genomic DNA; 4) delivery of genome editing vector into cells in culture; 5) some cells in culture successfully edit their genomes on-target, others edit their genomes at off-target sites, the majority of cells remain unedited. A more detailed version of the “Ideal CRISPR Genome Editing Process” is made available in Section 2.

A great many tools have been developed to bring about a wide variety of useful genomic interventions. What has been described can be regarded as a very generalized example of genome editing as a process. Variants of it have been widely adapted, forming a taxonomy of applications used in academic, commercial, and medical settings, where cells, plants, and animals can be genome edited with impressive outcomes. This taxonomy is described in Section 3.

## Why Edit Genomes?

Genome editing can be used experimentally to develop knowledge that proves that something is so (basic science), or that some exploitation or application of this knowledge is possible (translational science).

Typical investigations involving a genome editing procedure might see edited cells compared with unedited cells to draw conclusions through comparative genomics. Removing a gene may make its function more obvious by its absence, or the addition of new genes may boost an attribute and reveal the gene’s function. A non-exhaustive list of ways to leverage the experimental capabilities of genome editing includes: loss of function, gain of function, substitution of one function for another, interference of a function in the presence of a certain

chemical, activation of a function with a certain chemical, or saturation mutagenesis where successive mutations are tiled across a region to determine what parts of a gene are responsible for which function.

Genome editing has been well-supported by an environment with 50 years of historic innovation in biotechnology, and a healthy market that both supplies and demands new genome editing tools and capabilities in an increasingly commoditized manner.

In the commercial sector, genome editing capabilities and activities are used across the life science industry, biomedical sector, agricultural biotechnology industry, and other sectors, with genome editing being either directly sold as a product or service, or otherwise used to drive intellectual property generation and deliver new products and services that leveraged genome editing in their development. A number of high value technology development and commercial activities involving genome editing processes are ongoing today, including the development of research reagents, diagnostics, bioproduction platforms, trade in commercial cell lines, trade of edited organisms, and the development of human therapeutics.

Edited organisms can be sold directly for use in labs for experimental research, or for their improved traits. Organisms modified to express the genome editing nuclease Cas9, for instance, make it simpler for researchers to edit that organism's genes, generating value for stakeholders in the research tools and pharmaceutical markets. On the other hand, gene edited crops could be sold directly on the market, for their ability to grow faster in harsher conditions than unedited alternatives. In medicine the prospect of targeted genetic surgery to treat or cure genetic disease has drawn much interest from clinicians and biopharmaceutical companies alike. For instance, CRISPR has been used demonstrate the clinical potential of deactivating the viruses in pig cells that make them unsuitable for human transplant (Niu et al. 2017), or it could more generally be used as a tool to prove that a gene or particular mutation plays a particular role in a disease.

In certain commercial settings, the value found in the ability to quickly derive new strains of an organism is maintained across the value chain, including those who develop those strains (biotech companies), private enterprisers who deploy them (farmers and crop growers) and purchasers of resulting crops (distributors or consumers). This is not to mention the value that investors might derive from holding stock in relevant biotech companies, or that governments and societies might benefit from with a more secure and productive food supply, superior medicines, or other elements of the bioeconomy.

## **How is Genome Editing Different from Traditional Techniques?**

Whilst genome editing was quickly achieved in yeast and mice soon after the first successes in genetic engineering (Rothstein 1989; Thomas et al. 1986; Scherer & Davis 1979; Smithies et al. 1985), the technique was cumbersome and difficult to direct, as the technology and tools available at the time were not well suited to this task. Genomic integration often occurred at random sites throughout a genome, potentially disrupting other mission-critical sequences, with successful editing occurring at considerably low frequency amongst cultured cells. The majority of early genetic engineering projects instead relied on inserting DNA into organisms at non-genomic sites such that this additional DNA could function outside the context of genomic DNA. Inserted DNA functioned so long as the introduced DNA was able to avoid

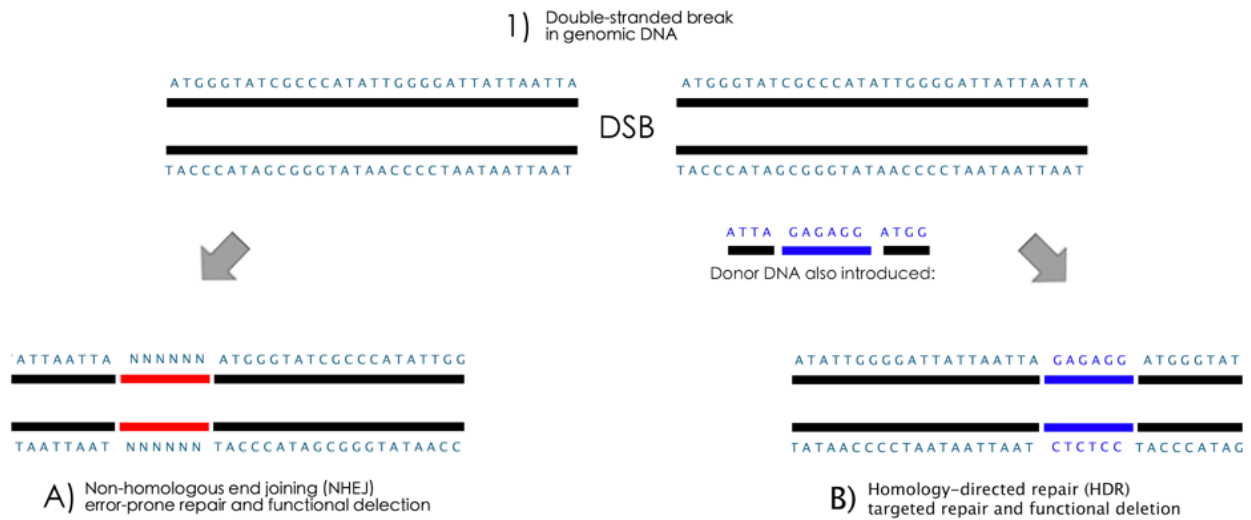
degradation or expulsion by the cell. For the most part, this extrachromosomal DNA approach was only practical in bacteria and other microbes, and while there was much success in microbial genetic engineering, progress in mammalian (and therefore human) cells was slow - as it was with many other non-bacterial organisms (Gaj et al. 2013). A specific, directable, and scalable way to integrate permanent edits to the genomic DNA any cell in any organism was needed.

## DNA Cleavage and Repair at a Glance

Genome editing tools typically introduce stable mutations at specific target sites by cutting both strands of a dsDNA molecule, which stimulates one of two classical repair mechanisms that naturally exist in many cells: non-homologous end joining and homology directed repair.

The non-homologous end joining (NHEJ) pathway repairs breaks in double stranded DNA by leveraging the sequences present in the overhangs of the cut. NHEJ usually results in an imprecise repair that preserves the overall integrity of the whole DNA molecule (the sequences upstream and downstream of the cut) but compromises the single genetic element bridging the cut by insertion or deletion mutations (InDels) at that site, likely leading to loss-of-function. As such, NHEJ can be exploited to functionally delete, or knockout, a particular gene.

The homology directed repair (HDR) pathway repairs breaks in double stranded DNA by leveraging DNA sequences with high similarity (homology) to the target site. In genome editing, an additional DNA *donor molecule* with high homology to sequences occurring upstream and downstream of the target site, is co-delivered to the cell alongside the GEV. The HDR repair mechanism incorporates the donor into the target site, and in doing so will also incorporate any other additional sequence information that is intentionally embedded in the donor. In this way, HDR can be used to precisely insert a novel gene or functional element, or otherwise overwrite a specific element. HDR may only occur during a particular phase of a cell's life cycle, and some cells which do not enter this phase are unable to be edited with this pathway. Timelines for these events vary depending on the organism, cell, GEV format and delivery technique used.



*Figure B. Schematic of two common genome editing mechanisms. A nuclease introduces a double-stranded break (1), which is repaired either through: A) Non-homologous end joining pathway which is an error prone repair mechanism that makes use of a cell's naturally occurring DNA repair mechanism, and results in a truncated protein that is non-functional or; B) Homology directed repair (HDR) pathway, which makes use of an exogenous DNA donor that is included in the genome editing payload alongside the GEV, causing the cell to integrate it at the target site.*

It is often overlooked that genome editing nucleases and programmable elements are only indirectly responsible for a genome editing event - rather it is the cell's repair pathways that complete an editing event stimulated by the initial cut. The NHEJ or HDR pathways are not found in all organisms - in some cases both are present, in others HDR is absent. In some cases, the genetic machinery required to achieve these repairs may need to be added to the cell somehow in order to remedy this issue, but this does add extra complexity to the genome editing procedure overall.

## Genome Editing Vectors and Delivery Options at a Glance

Programmable nucleases, and their customized elements (binding domains) must all be introduced to a target cell to conduct genome editing. As DNA codes for RNA, which in turn codes for protein (known as the central dogma of biology), GEVs can come in any of these three formats, or can otherwise be housed in larger more complex molecular vehicles, such as viruses, bacteria or engineered nanomaterials. The most important and commonly used GEVs for genome editing tools are as follows:

- **DNA plasmids:** these are circular molecules of non-genomic DNA that contain DNA-based instructions. Plasmids were initially derived from bacteria, which use them to transfer genes between individuals, and have been commonly used as “USB sticks” to shuttle code into cells. In the context of genome editing, after delivery into a cell, the cell will respond by translating the plasmid DNA into RNA, and then into the relevant proteins, including

nucleases and binding domains. Plasmids may serve as GEVs in of themselves, or otherwise hold further instructions to package CRISPR elements into a more complex GEV (such as a virus). They can be preserved and replicated with ease.

- **Pre-transcribed RNA transcripts:** these nucleic acids act much like DNA, but can directly code for a protein. After delivery into a cell, the cell will respond by directly translating the RNA into relevant proteins, which will take less time than a DNA plasmid-based approach, as the first DNA-to-RNA step is unnecessary.
- **Viral particles:** these are first synthesized as DNA which is booted up in a dedicated bioproduction cell line that allows the designed viruses to be farmed in high number. The appropriate genome editing elements are engineered into this virus during this initial bioproduction phase. Cells that are to be edited are infected with the resulting virus, which in the case of a *lentivirus*, will integrate elements of its viral genome into the host genome, at a random site. At this point, the host genome which now contains the genome editing elements in DNA form, will be expressed as RNA, and converted into functional nuclease proteins and binding domains. Viruses act as delivery vectors in this way and can express species-specificity. Many different viruses are available to be used as delivery vectors for different organisms, some of them integrative, others non-integrative.
- **Synthetic ribonucleoproteins (RNP):** these are associations of RNA and DNA-binding proteins that essentially form pre-complexed genome editing vectors that are immediately functional in a cell. This method is the fastest-acting, as there is no need for DNA to be transcribed into RNA, and into protein in the cell.

*Genetic transformation* is a term that collectively refers to the successful physical delivery of the GEV as well as a successful genetic alteration. Transformation methods include: heat-shock transformation, chemical transformation, viral transduction, biolistics (gene guns and nanoparticle delivery), bacterially-mediated transformation, and exposure to silica whiskers laden with DNA, some of which are discussed in more detail in section 3. It should be noted that the terms relating to GEV format and transformation technique may be used interchangeably. Moreover, some GEVs are better suited to a particular type of genome editing tool and a particular organism or cell type, whereas other GEVs cannot be delivered with certain transformation methods. The information here is simply meant to illustrate the variety of options available, and ample literature describes additional options not mentioned here.

It is noteworthy that a good deal of innovation in delivery approaches has occurred in recent years in response to CRISPR, which due to its small size, works well with a wide variety of GEV and transformation approaches. The efficiency of editing can vary wildly based on the cell line, organism and delivery method / GEV format used. Plasmid GEVs are generally used to deliver each of the three main genome editing tools, and going forward, are used as the “standard” GEV in this paper. Figure C provides further explanation of the precise mechanisms used by some GEVs to boot up the necessary biological machinery in a living cell.

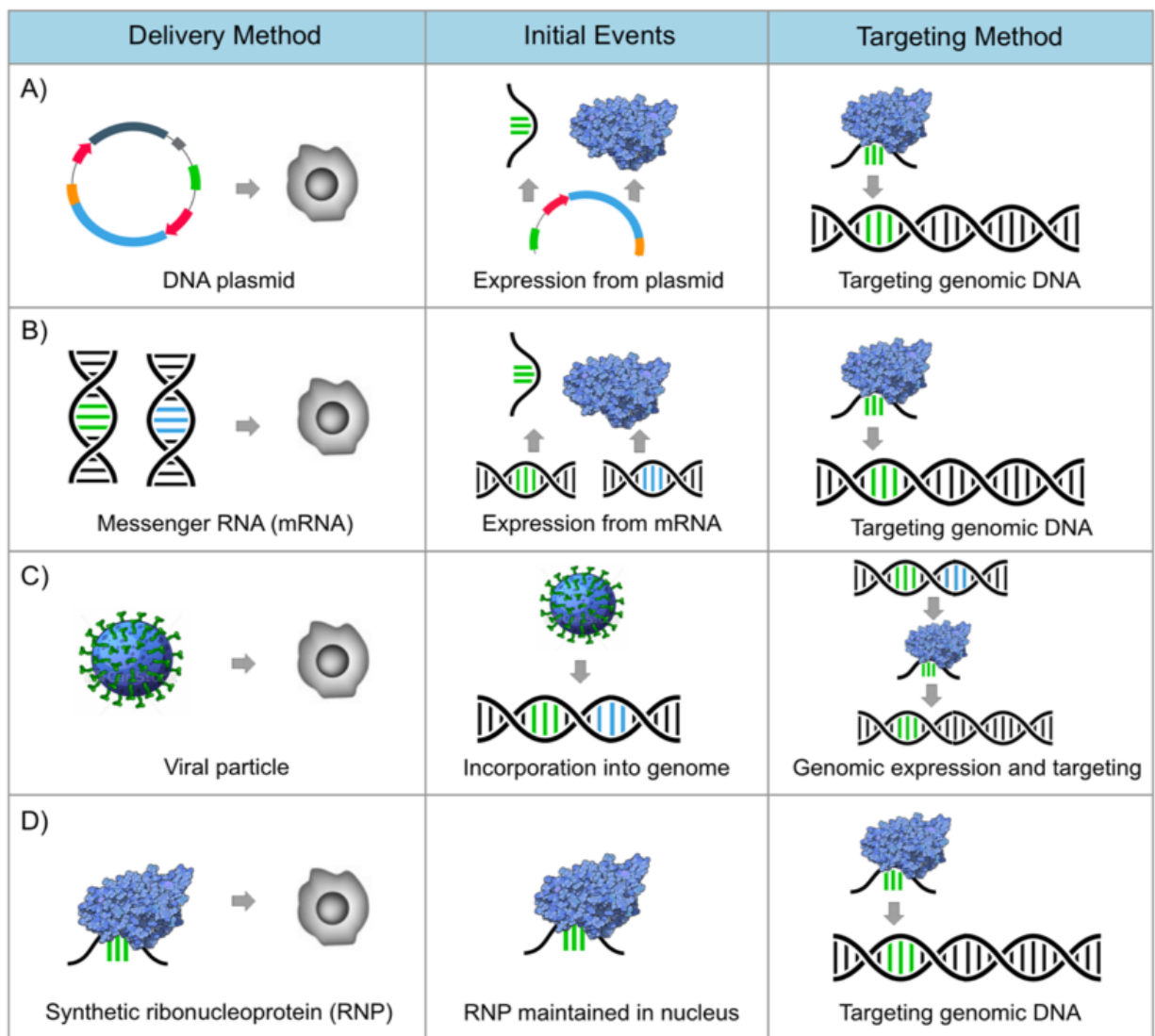


Figure C. Schematic of different GEV delivery methods, showing: A) DNA plasmid introduced to a cell such that a nuclease and binding domain are expressed before complexing together to cut a target genomic DNA sequence. B) Two separate messenger RNA molecules, together coding for a nuclease and a programmable domain. C) A viral particle containing DNA that codes for a nuclease and programmable domain, which is incorporated into a host cell's genome, subsequently expressed by the target cell to cut its own DNA. D) Synthetic RNP delivered into a cell whereupon it can immediately cut a target site before being degraded.

## Genome Editing Tools at a Glance

Three genome editing tools have received much focus in the last decade: Zinc Finger Nucleases (ZFNs), Transcription Activator Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Palindromic Repeat (CRISPR) systems. Whilst a number of other families

of genome editing tool exist, such as the engineered meganucleases, we focus here on the most well-known tools in order to provide context for the success of CRISPR compared with historically-used well-established techniques.

## Zinc Finger Nucleases (ZFNs)

Zinc-finger nucleases (ZFNs), introduced in 1996 (Kim et al. 1996), were derived from a decade of basic research on the nature of zinc-binding proteins in model organisms such as *Xenopus frogs* (Klug et al. 1986; Klug 2010). ZFNs make use of a pair of effector molecules, and in each of the doublets, there is a separate DNA binding domain attached to a nuclease domain. Binding domains are built from arrays of amino acids forming a zinc finger. Each zinc finger can bind to anywhere between 9 and 18 base pairs in a sequence-specific manner (Liu et al. 1997), and the modules can be chained together in order to target a variety of complex sequences. The nuclease component is a FokI protein that is able to cut a single strand of DNA. Two ZFN doublets bind to each strand in a manner that aligns two FokI domains over the same position, as shown in in Figure D.

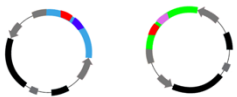
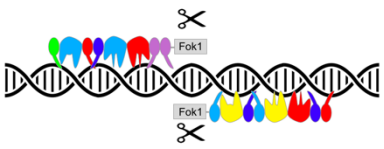
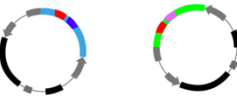
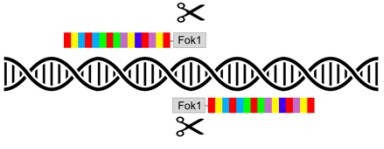

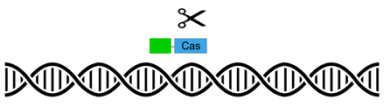
Delivery Complexity	Targeting Schematic	Summary
<p>A) Zinc Finger Nuclease</p> 		<p>Targeting - almost unlimited Specificity - high Predictive challenge - high Accuracy - high Engineering - very hard</p>
<p>B) TAL Effector Nuclease</p> 		<p>Targeting - some limitations Specificity - high Predictive challenge - medium Accuracy - high Engineering - hard</p>
<p>C) CRISPR nuclease</p> 		<p>Targeting - some limitations Specificity - medium Predictive challenge - easy Accuracy - high Engineering - very easy</p>

Figure D - A) Classical Zinc Finger configuration, showing doublet FokI nucleases overlapping at a target site, each attached to a unique multi-part binding domain. B) The same for TALEN configuration. C) Classical CRISPR-Cas9 configuration, showing a single Cas9 enzyme attached to a single guide RNA. Notice the relative simplicity of the CRISPR GEV format versus Zinc Finger and TALEN libraries.

Individual zinc fingers can overlap and alter one another's sequence-binding specificity making it difficult to predict the emergent binding specificity caused by this interference. As such, developing ZFNs for arbitrary sequences is a fairly empirical process requiring laborious screening and optimization (Boch 2011), with unexpectedly high failure rates despite great efforts to identify interference and binding patterns (Ramirez et al. 2008), and develop relevant protocols (Wright et al. 2006). ZFNs are largely delivered as plasmids, but ZFN projects are



heavyweight, requiring the delivery of multiple plasmids coding for different zinc finger domains, in what is termed a library. Use of this library could require many rounds of transformation and delivery over several months, in order to produce a unique zinc finger protein for one experiment. Nevertheless, the resulting proteins would have very few off-target effects due to the length of the targeting arrays.

### Transcription activator-like effector nucleases (TALENs)

Transcription activator-like effector nucleases (TALENs), were introduced in 2009, and derived from the pathogenic bacterium *Xanthomonas* (Boch et al. 2009; Morbitzer et al. 2010; Christian et al. 2010). Like ZFNs, they make use of a paired doublet of two-part protein complexes (Figure D). The first part of each doublet is a *DNA-binding domain* built from arrays of 33 to 35 amino acid modules, each of which targets a single nucleotide in the target site. The modules are chained together in a particular order to target a variety of sequences. The second part of the doublet is *FokI endonuclease protein* that can cleave DNA. By assembling two complete doublets on either strand of a target site, researchers can selectively target a sequence and precisely introduce a double strand break (DSB) at the point where the FokIs overlap and each cut a single strand.

TALENs are typically delivered as plasmids. Like ZFNs, these projects are fairly heavyweight, requiring extensive work to assemble the two chains of up to 35 amino acids into a protein binding domain. TALENs may, however, be considered a step change as compared with ZFNs as their design is more straightforward - a single TALEN sub-unit is able to recognize a single nucleotide, reducing the potential for emergent interference and unpredictable binding, thereby enabling more effective prediction, and reducing the experimental overheads required to confidently target an arbitrary sequence (Zhang et al. 2011).

Every TALEN project requires specific design and manufacture, which could take weeks or months (Anon n.d.; Moore et al. 2014), and came with some targeting limitations (Gaj et al. 2013), but like ZFNs there were few off-targets due to the length of the targeting arrays (Veres et al. 2014).

### CRISPR and RNA-guided Endonucleases (RGENs)

The introduction of CRISPR as a genome engineering technique occurred between 2012 and 2013 thanks to recognition that a CRISPR-based immune system in bacteria could be repurposed and repackaged into other organisms to introduce double-stranded breaks at a target site (Cong et al. 2013; Jinek et al. 2012a; Mojica et al. 1993). The field has rapidly augmented CRISPR systems into a suite of RNA-guided endonuclease (RGEN) tools that are easier and cheaper to design and use, and are more versatile than ZFNs and TALENs. The nomenclature tends to reference a CRISPR/protein version being used, for instance CRISPR/Cas9, but one should understand that CRISPR is a fairly loose term encompassing a wide variety of RGEN variants that exhibit the same general RGEN behavior and anatomy, with characteristic PAM requirements, preferences for sequence binding, and expected cleavage outcome tied to a computational method that works best with more characterization backing it up.

CRISPR RGENs have a constant nuclease called a *CRISPR-associated* or *Cas* protein, with multiple functions, one of which is to bind to dsDNA, unwind it and introduce a double-stranded break at a target site. A single RNA-based binding domain acts as the programmable

element, bringing significant advantages as compared to ZFNs and TALENs. Rather than require a customized protein to be engineered in a stepwise fashion, the guide RNA can easily be introduced into a single unloaded CRISPR plasmid GEV template in a process known as *molecular cloning* which assembles an experiment-ready construct. Whilst ZFN and TALEN plasmid GEVs are also cloned, the operation can require many months to assemble and test the plasmids required in a trial and error genome editing project. As for CRISPR (see figure D), the GEV plasmid cloning steps are reduced, requiring a single short custom oligonucleotide to be cloned into a more comfortably standardized plasmid template. Another feature of CRISPR is that it is well suited to a variety of GEVs beyond plasmids, including mRNA, viruses, and synthetic RNP. Loaded CRISPR GEVs can easily be delivered to a cell, and whilst they differ in format, the generally all “boot up” a payload that complexes together as shown in Figure E.

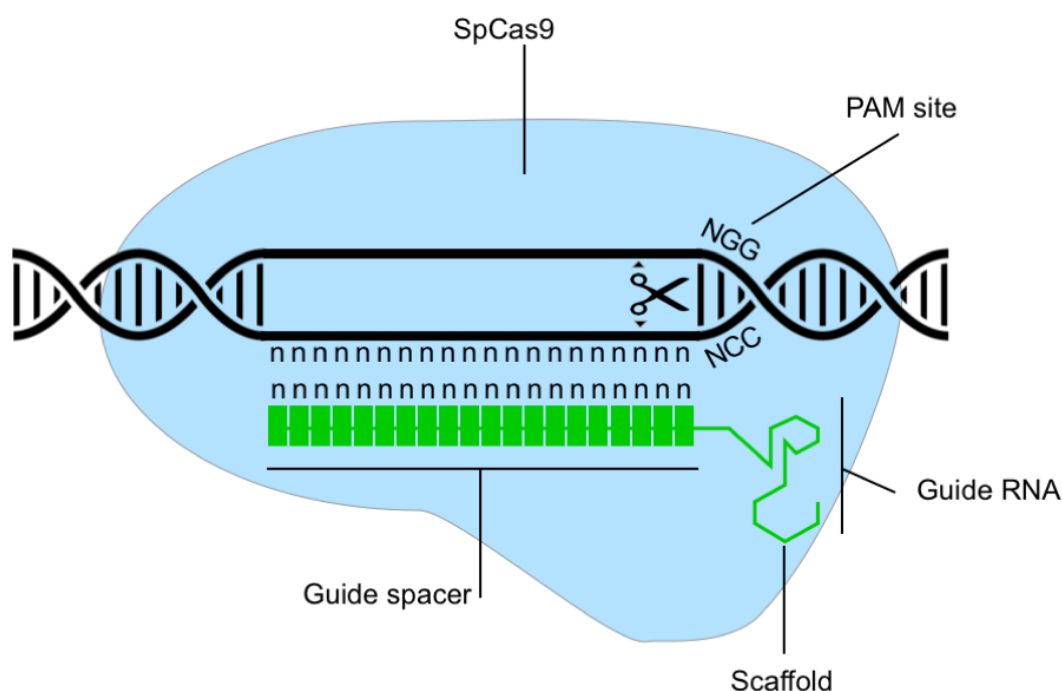


Figure E. Schematic of SpCas9 and guide RNA complexed with an unwound dsDNA molecule. The SpCas9 nuclease unwinds the genomic DNA at a PAM site with the NGG sequence, and cuts at the site where the guide RNA spacer matches one of the strands of dsDNA.

The most widely used and well-studied CRISPR RGEN tool uses a Cas protein derived from the bacterium *Streptococcus pyogenes*, commonly referred to as SpCas9. An important limit on SpCas9’s utility is a requirement for a specific sequence to occur in proximity to the targeted cut site called a protospacer adjacent motif (PAM) site, with an NGG sequence. In other words, wherever there is a genomic DNA site with a sequence ‘GG’ the SpCas9 nuclease is able to unwind the genome and potentially introduce a DSB.

As additional bacterial immune systems have been characterized and harnessed, a variety of Cas proteins have emerged for use in CRISPR genome editing. *Staphylococcus aureus* Cas9

(SaCas9) (Ran et al. 2015), for instance, is a nuclease protein that is smaller than *SpCas9* meaning it can more easily be housed in a viral genome editing vector. CRISPR RGENs from *Prevotella* and *Francisella* (largely known as Cpf1 but recently classified as Cas12 (Chen et al. 2017) have properties that make them better-suited for use in regions that *SpCas9* cannot cleave, and for cutting the genome in a manner that makes it useful in non-dividing cells (Zetsche *et al.*, 2015).

*SpCas9* is the most widely cited and used CRISPR RGEN, and has been found to be a more efficient genome editing tool than TALENs for NHEJ knockout but not HDR knock-in by at least one group (He et al. 2016). *SpCas9* guide sequences are shorter than that of a ZFN or TALEN (a 20-mer<sup>1</sup> CRISPR guide sequence compared with the 36mer ZFN and 40mer TALEN<sup>2</sup>) (Koo et al. 2015), and while ZFNs and TALENs can conceivably be engineered to target any sequence in the human genome, an average human genome is reasonably expected to have around 161,000,000 NGG PAM sites, meaning that a potential cut site could be found every 42 bases. It should be noted that not all 161,000,000 CRISPR-nuclease guides will cut with the same performance - some will have outstanding performance, whilst many will not cut well. Additionally, CRISPR guides that perform well in one nuclease would not have the same performance characteristics for another (if they work at all).

A large number of computational tools are available to predict off-target and on-target binding of guides (Ran et al. 2013; Shalem et al. 2014), and the challenge of finding the right guide for the job is simpler than ever before. A variety of nuclease proteins have different PAM sites, providing alternatives if one guide-nuclease pair is not suitable for a target gene, and the increasing number of RGEN systems that are available adds further options. While any individual guide could be worse than the average ZFN/TALEN binding domain, the large number of options and effective predictive tools available makes it simpler to find suitable CRISPR target sites than for older generations of genome editing tool. Although there are some limitations in computational design approaches, especially for newer Cas proteins which are yet to be well characterized, work continues to expand the PAM options of the already-well-characterized *SpCas9* by evolving it to further accept non-canonical PAM recognition sequences (Wrighton 2018), as well as to engineer higher fidelity RGENs with improved guide-genomic DNA recognition (Slaymaker et al. 2016; Kleinstiver et al. 2016).

CRISPR systems have, and continue to be, adapted beyond the scope of genome editing, into a broader genome *engineering* platform, through which genomic DNA can be manipulated beyond the scope of NHEJ-based knockout or HDR-based knock-in. A final feature of CRISPR RGEN systems is the requirement for one nuclease protein, rather than the doublet *FokI*s found in older generation technologies.

Further user-innovation in the CRISPR RGEN field is described in Section 2 with great attention paid to the biological and economic drivers of its uptake by the life sciences research community and biotech industry.

---

<sup>1</sup> An oligonucleotide with a sequence length of 20

<sup>2</sup> Actual oligo synthesis lengths may typically requires slightly longer oligos, depending on the tool and the GEV format

## Limitations of Genome Editing Tools

### Prediction of Editing

Biologists are limited in their time and financial resources, and are thus limited in the number of genome editing runs they can perform for any given experiment or procedure before exceeding a resource budget. In genome editing, the prediction of targeting helps biologists to identify the programmable domain sequences most likely to yield desirable experimental outcomes, and fit their experimental runs within this budget. Predictive power is determined by the availability of bioinformatic tools trained on a dataset of historical genome editing outcomes for a particular tool in a particular context. For instance, a dataset that lists the efficiency of InDel mutation formation for 2000 different CRISPR guide RNA sequences in a mouse genome can train a computational algorithm to score an arbitrary guide RNA sequence for likely activity in a similar mammalian genome (a rat or human). Algorithms trained on more comprehensive datasets with more explicit examples of editing outcomes allow more sophisticated predictions to be made than otherwise allowed by smaller datasets. In addition to these algorithms biologists also require at a minimum, a digital representation of a target gene, and ideally a reference genome of an organism, or better yet, the actual genome sequence of the individual organism they are working with. The availability of these computational tools and sequence datasets can serve as a limiting factor on the use of different genome editing systems, and whilst genome editing can still technically be achieved without some (or all) of them, editing procedures become increasingly complicated and error-prone as these essentials are removed from the equation.

### Off-target Challenges

Computational tools or scoring algorithms can be used to predict both desirable editing outcomes, and also flag up the risk of potential off-target (undesirable) editing events where a binding element guides a nuclease to cut at an unintended site. Inadvertently introduced off-target edits can disrupt normal physiological function of a cell's genes or metabolic pathways, degrading or debilitating its ability to function in certain conditions, which may not be obvious to a biologist in the first instance (Pattanayak et al. 2011; Fu et al. 2013). Off-target effects can have serious implications for experimental accuracy in the lab, and clinically significant side-effects when they occur in medical applications of genome editing. Off-target effects are typically caused by sequence similarities between the on-target site, and partially or exactly alike off-target sites elsewhere in the genome. Potential off-target sites can be predicted using computational tools, but the validity of the conclusions drawn are only as good as the sequence information used as the basis of that prediction (and the sophistication of the training data made available to the predictive tool).

A common misconception of off-targets is that they serve as a barrier for successful genome editing, but the nature and impact of this barrier will depend on the accuracy of editing required for the application in question. For a demonstration that a genome editing tool can work in a novel organism, off-targets would be acceptable, presenting no barrier in the context of such a project, whereas for a clinical intervention any off-targets in a patient's cells should be avoided at all costs. Indeed, without sufficient proof that off-targets are avoided, a CRISPR therapeutic is unlikely to overcome the barriers of many clinical regulatory agencies' standards of safety.

## Quantifying and Validating Editing

In order to validate conclusions and provide confidence in results, it is important to quantify genome editing efficiency at target and off-target locations. Quantification and validation may occur in a targeted manner, in which predictive tools are used to determine and sequence the most likely off-target editing sites, a method which relies on the strength of the predictive tool, sequence information available, and resolution of sequencing instrumentation and informatics. Alternatively, quantification and validation may occur in an untargeted or whole-genome sequencing approach, where the genome (of a population of cells) is sequenced before and after the editing run. A number of assays and approaches have been described for detecting editing events in this manner (Tsai et al. 2015; Gabriel et al. 2011; Kim et al. 2015; Frock et al. 2015). The most salient challenge for detecting and quantifying genome editing events is in the identification of extremely low-frequency off-target events that go unpredicted, likely lying beyond the scope of training data or the resolution of sequencing instrumentation and analysis software packages. It has been noted that exponential growth in sequencing capability has not translated into parallel increases in the sensitivity of these technologies to detecting rare mutations, namely due to noise arising from the error rate of next-generation sequencing platforms (Tsai 2017), as well as the fact that off-target edits may not have occurred in the cells being sampled, rather occurring in cells that are unsampled (there being a practical limit on the amount of tissue one can reasonably assess from any organism). With current technologies it would be fair to say that there is a limit beyond which editing events become challenging to detect and quantify. Whilst great debate has occurred in recent years around the impact of the off-target barrier in CRISPR systems (Schaefer et al. 2017; Wilson et al. 2018; Lareau et al. 2018; Nutter et al. 2018; Kim et al. 2018; Lescarbeau et al. 2018; Schaefer et al. 2018), at least one group has announced a capability to edit without any unexpected off-target editing events (Iyer et al. 2018).

## Gene Copies and Mosaicism

Organisms can have more than one copy of a gene, and whilst the sequence of this gene can be targeted, not all copies may be successfully targeted and modified in the same way. This can give rise to mosaic cell populations or whole organisms, in which cells have different genotypes derived from the same genome editing procedure IE different cells' genomes are disrupted in different ways, potentially leading to different rates of gene expression between cells or tissues. Mosaicism can imply partial knockdown of gene expression rather than complete gene knockout, and is especially important in cases where a gene dosage phenomenon is observed and the number of gene copies have a proportional effect on the phenotype that is observed. Mosaic organisms in the first generation can be bred together in order to segregate and isolate animals or plants that carry the desired mutations in the second generation. The production of cells or organisms with multiple genotypes is variably considered a problem, generally depending on the context and project objective. For instance, mosaic cell therapies used in patients would be intolerable for almost every clinical intervention as each cell could perform differently. Mosaics in animal editing could be better tolerated as breeding another generation of mice to obtain the desired mutant lineage might fit comfortably within the project timeline.

## Somatic and Germline Editing

Mutations are typically described as belonging to one of two classes: germline mutation and somatic mutation. In germline mutation events, a change to the genome occurs in a gamete cell, namely unfertilized sperm and eggs or a fertilized single cell embryo, and the mutation is said to be heritable from one generation to the next. In somatic mutations, genomic changes occur in a body cell, and cannot be passed on to the next generation (the exception is if an organism were cloned from the somatic cell, such that a second cloned individual would also obtain the same mutation). When it comes to genome editing, there is stark difference between how somatic and germline editing is dealt with. Germline editing is typically more difficult to achieve as it requires an operator to have capabilities to deal with gamete cell biology and fertilisation processes using precious and limited cells, whereas somatic cells can typically be obtained and handled more easily in bulk culture using simpler cell biology capabilities. However, those who succeed with germline editing procedures will produce a self-sustaining lineage of a whole organism that can pass on the desired edit to its offspring. In germline editing it is necessary to deliver genome editing payloads at the single cell stage of embryogenesis such that all cells that derive from this embryo in the adult, including the sperm and eggs, have the same genotype, and are not mosaics. If multicellular embryos are edited, mosaicism may be observed across different tissues in the adult, and if the gonadal tissues do not carry the mutation, the mutation is not heritable to their offspring.

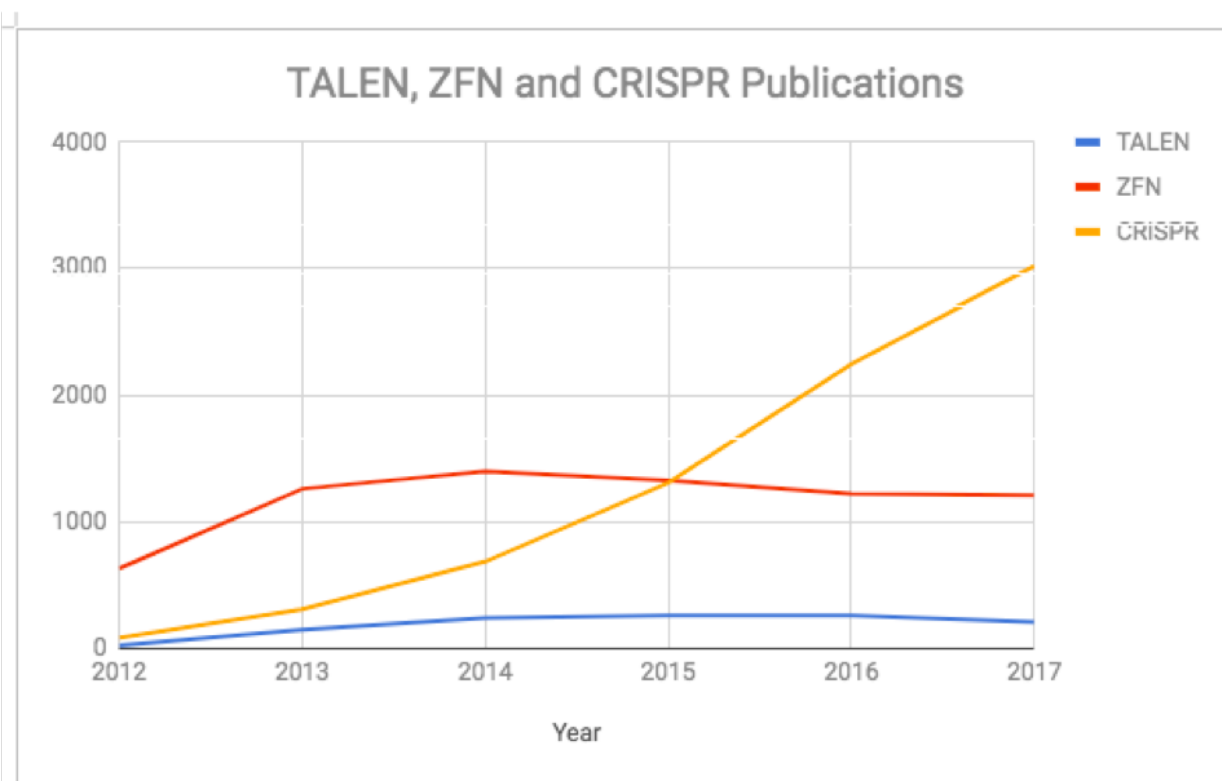
## SECTION 2: DRIVERS OF CRISPR ADOPTION

Recent uptake in CRISPR by life science researchers has been incredibly rapid due to its biological flexibility, wide utility in biological engineering applications, and a number of favorable economic factors. This section explains how CRISPR has come to be so widely used amongst molecular biologists.

### The Wide Adoption of CRISPR

Between June 2012, the date of the first paper describing the use of CRISPR as a genome editing tool (Jinek et al. 2012b), and the end of 2017, there have been at least 142 press releases regarding new CRISPR products and discoveries, and 8,074 CRISPR papers published by more than 54,133 authors and co-authors (Thompson & Zyontz 2017), an average of 125 papers per month. CRISPR has now emerged as the *de facto* genome editing tool of choice, perhaps best

illustrated by the waning ratio of publications listing TALEN or ZFNs compared to CRISPR, as shown in Figure F.



*Figure F - total papers published relating to Zinc Fingers, TALENs and CRISPR, between 2012 and 2017. Data compiled from Google and PubMed.*

In the lab, CRISPR offers the molecular biologist a number of biological advantages over ZFN and TALEN in terms of accessibility, scalability, economics and infrastructure. In short these advantages include the: simplicity of target design and GEV construction; the relative ease of use and control of CRISPR systems; high rate of user innovation and use case flexibility; and wide support across versatile organisms.

An additional set of non-biological and systems-level drivers for CRISPR adoption include: permissive IP structures for research use and; a unique history of permissive technology distribution and enabling infrastructure from day 1.

Together, these drivers have not only met biologists' requirements for an ideal tool, but enabled this primed market to rapidly obtain access to the explosively popular CRISPR technique.

## **Biological Drivers of CRISPR Adoption**

### **Simplicity of Design and GEV Construction**

CRISPR functional GEVs can be relatively easily assembled, which helped to transition genome editing from being a complex custom protein engineering and cloning-intensive exercise into a simpler project with a single round of cloning.

Simply put, the labour required to design and assemble first and second generation genome editing tools such as ZFNs and TALENs placed limits on the practical utility of genome editing processes. ZFN engineering has an onerous requirement to select and screen new zinc finger proteins for their affinity to particular genome sequences (Durai et al. 2005), adding significant complexity to their discovery and general use. As for TALENs, although their predictive design was easier, the assembly of custom binding domains remained troublesome given the need to engage in sequential cloning operations (Christian et al. 2010), despite efforts to create a more modular design approach using simpler cloning methods (Cermak et al. 2011). These issues were comfortably addressed by CRISPR, which removed a key challenge of generating a custom DNA binding domain in a lengthy and step-wise manner. Arbitrary CRISPR guides can be reliably designed with predictive methods, supported by numerous software tools, which allow researchers to easily select the best-performing guides with improved confidence. CRISPR has thus become attractive for many experiments as a lightweight tool that can help answer many questions in biology.

In practice, the simple construction mechanics mean that cell line engineering, which used to take up to 90 days (or more with design optimization), can now be achieved in as little as two weeks by more junior staff without complex protein engineering experience. Simple construction mechanics also make it practical for larger genome editing experiments to be conducted as thousands of GEVs can be produced at scale with relative ease.

### **Ease of Scale and Control**

CRISPR scales well as it allows researchers to conduct repeated genome editing experiments or procedures at relatively low financial costs and time burdens. The simplicity and affordability of CRISPR makes it suitable for scaled up applications in a laboratory. Further, many laboratories were capable of taking advantage of the CRISPR capability without investing in additional equipment.

CRISPR also has elegant experimental scalability at the molecular level. Due to the simple nature of the basic CRISPR plasmid and small size of several of the molecular components involved, RGEN packages can be assembled to have multiple functions. A single CRISPR GEV can house >1 guide RNA, allowing it to target multiple copies of a gene (or similar variants) at once. A noted record for this “multiplex genome editing” is the successful targeting of 62 highly similar sequences across a genome, edited by two guides on the same GEV (Yang et al. 2015). Further, the ease of construction allows biologists to readily assemble libraries, using simple software approaches, to conduct more advanced applications.

More generally the small size of CRISPR systems has supported healthy innovation in delivery approach and GEV format, each providing different layers of control. CRISPR can be introduced in multiple delivery formats (described in Section 1), and the expression of nucleases



can be readily tuned with a wide range of inducible expression strategies, some of which may be fairly simple, with one chemical driving expression of a single nuclease, or more complex with multimodal chemicals used to variably alter the conditional expression of CRISPR devices to target multiple genomic loci (Lu *et al.* 2018).

Easy and cheap scalability, plus fine-tuned control of molecular events, allows reproducible data to be obtained comfortably, further cementing the utility and pull of the CRISPR tool and RGEN approach.

### **User Innovation: From Simple Tool to Flexible Platform**

Since the development of the first SpCas9 approaches, CRISPR and RGEN systems have become a platform technology upon which much ingenuity has been applied. Today a growing palette of advanced RGEN and RGEN-like tools with interesting properties has been made available by numerous user-innovators and researchers. Notable examples include:

- The “nickase technique” that makes use of paired CRISPR RGENs that are each engineered to nick a single strand (rather than a single nuclease cleaving both strands) to reduce off-target risk (Shen *et al.*, 2014).
- CRISPR interference (CRISPRi), which involves the use of a deactivated Cas9 (dCas9) fused to a domain that blocks DNA transcription machinery from binding to a target, effectively interfering with the expression of the gene in a time-limited fashion (Larson *et al.*, 2013).
- CRISPR activation (CRISPRa) which uses a transcriptional activator to selectively express the target gene (Gilbert *et al.*, 2014).

Such techniques rely on CRISPR’s elegant ability to bring together DNA, RNA and protein with fairly simple mechanics, and achieve complex outcomes. By fusing multiple functional elements to an RGEN scaffold, there is impressive scope to extend the capability of CRISPR far beyond its role as a simple genome editing tool used for knockout and insertion of DNA. Rather, CRISPR RGENs become a platform technology that can support a wide range of genomic interventions at almost any target site. The magnitude of this flexibility cannot be overstated, and continued innovation in this area will potentially require further redefinition of the very concept and scope of genome editing in decades to come.

User innovations beyond editing provide even more reasons for researchers seeking advanced capabilities to experiment with CRISPR. This is just a tiny set of examples, but the outlook for CRISPR’s continued development is positive as new Cas variants continue to be developed, and new RGEN proteins are discovered and made available. Indeed, the technology has attracted compelling metaphors including comparisons with both a Swiss army knife and molecular find-and-replace tool, and it is clear that the biological properties and modular nature of CRISPR make it an ideal platform for continuous user innovation and flexibility. What is unclear is where these capabilities will run up against technical limits, and when user-innovation will top out, if ever.

## Wide Organism Versatility

CRISPR has drawn attention from groups working in all sectors of biology because of its purported high performance in numerous organisms. CRISPR was quickly applied to many animal model species with great success. Key applications and papers included the first CRISPR genome engineering of mice (Wang et al., 2013), zebrafish (Jao, Wente, & Chen, 2013), fruit flies (Bassett, Tibbit, Ponting, & Liu, 2013), nematode worms (Friedland et al., 2013), rats (Hu et al., 2013), frogs (Nakayama et al., 2013), and monkeys (Y. Niu et al., 2014).

In plants, CRISPR's broad translational applications in agricultural research encouraged its successful use in crops and their model organisms. In 2013 five independent groups reported the successful application of CRISPR genome editing in classic plant models *Arabidopsis* (J. F. Li et al., 2013), tobacco (J. F. Li et al., 2013) and rice (Miao et al., 2013). Genome editing techniques were also quickly recognized by the agricultural industry, with DuPont Pioneer publishing studies in 2015 on CRISPR modification of soybean (Z. Li et al., 2015) and maize (Svitashev et al., 2015). In 2016, they used CRISPR to generate maize which offered high yield in strenuous drought conditions (Shi et al., 2017).

The technology also works well in other eukaryotes of interest, including medically relevant species such as *Plasmodium falciparum*, a malarial protist parasite (Wagner, Platt, Goldfless, Zhang, & Niles, 2014), and numerous non-model plant and animal species.

While ZFNs and TALENs had also been shown to work in various plants and animals (Doyon et al. 2008; Sander et al. 2011; Tesson et al. 2011; Geurts et al. 2009; Bibikova et al. 2002; Maduro 2006; Young et al. 2011; Carlson et al. 2012; Wood et al. 2011; Zhang et al. 2010; Li et al. 2012), it is the relative ease of GEV construction and other useful attributes that made CRISPR a top-contender for labs that historically sought a low-cost tool validated to work in their organism of choice. In other words, the opportunity cost for a lab to begin experimentation with CRISPR is low as compared to that of TALEN/ZFN because other labs have effectively derisked the chances of failure. The steady stream of publications explaining precisely how to adapt appropriate formats of CRISPR GEV to suit a widening range of edited organisms further cemented the tool's expansion. Further user-innovations made CRISPR even more attractive to researchers that typically see advanced capabilities reserved for humans and model organisms, rather than more esoteric organisms found in non-medical life sciences.

## Non-Biological Drivers of CRISPR Adoption

### Permissive Intellectual Property for Research Use

Despite legal challenges surrounding a complex patent litigation case to determine the ownership and rights of CRISPR across the globe (Sherkow 2018), academic and non-profit research has benefited from a permissive attitude to CRISPR use and licensing. In particular, a decision by the Broad Institute to make their fundamental intellectual property (IP) for CRISPR/Cas9 freely available for research use in academic and non-profit settings (Anon 2016), has opened the door to technology access by many users. Furthermore, non-profits, academic institutions and government agencies are permitted to transfer CRISPR materials to other nonprofits without the need to receive written licenses to do so. Put simply, academic and non-profit researchers do not have to sign up to onerous legal terms, high costs, or other administrative processes in order to

begin experimenting with CRISPR/Cas9 in the lab themselves, nor do they need to obtain permission to build productive communities around their research. Indeed, the actions of the Broad Institute set a high bar that other CRISPR IP developers must meet in order to generate sufficient attention, capture new users, and generate publications around novel IP.

The licensing of fundamental IP has also been fairly permissive in terms of commercialisation, with the Broad Institute pledging to non-exclusively license its IP to companies wishing to sell CRISPR tools, reagents, and services for basic and translational research. This move supported the development of a healthy product- and service-provider ecosystem, as numerous companies quickly moved into the space to offer products for a wide variety of genome editing applications, often competing with one another on a product- and cost-basis to drive down costs for the end-user. In comparison, the holders of fundamental ZFN and TALEN IP adopted a more restrictive approach to licensing and commercialization, with exclusive licensing deals seeing a handful of companies obtain monopolies on early generation genome editing technologies, leading to high costs per experiment and per kit. This IP strategy likely contributed to the eclipse of older tools by next-generation alternatives when they arrived on the scene.

### **Addgene's Wetware-on-Release Model**

The progression of CRISPR as a platform technology that works across multiple organisms went hand-in-hand with the development of new plasmids, which carry the instructions necessary for CRISPR and CRISPR-expression elements to be booted up in a cell. Addgene, a Boston, MA non-profit, is the most-used source for acquiring CRISPR-Cas9 capabilities (Thompson & Zyontz 2017) and has played a profound role in enabling the widespread use of the technology through its plasmid distribution service. Addgene was set up in 2004 to reduce the burden that laboratories faced in directly sending plasmids to those requesting them once their landmark publications (and novel plasmid sequences) were published (Joung et al. 2015). Addgene acts as a maintainer and broker of plasmids, collecting them from depositors, and supplying them to new users that are willing to reference the original publication and support by Addgene.

In terms of CRISPR, Addgene's service allows requesting scientists to avoid the need to redesign and test a novel plasmid themselves. Instead researchers can build upon pre-existing work from one author's publications by obtaining the plasmid from Addgene at the low cost of \$65 USD per plasmid requested.

With the Addgene CRISPR distribution model, researchers are still able to "brew their own" CRISPR constructs by augmenting requested plasmid chassis with a few oligos to assemble a new set of guides and novel GEVs. In this way Addgene supports CRISPR users by reducing a significant element of the work needed to become experiment-ready (the design and synthesis of the initial plasmid), and ensures that the costs are no more than the \$65 plus required oligos and reagents. Moreover, requesting scientists face almost no hurdle in paperwork as Addgene makes use of a standardized Materials Transfer Agreement (MTA) that has been widely pre-approved by institutions all over the world.

From the beginning, expert labs developing CRISPR itself deposited their plasmids with Addgene in advance of major publications. In this way, CRISPR plasmid requests could be made on the same day that a new paper and protocol was published, and requesting labs could receive the plasmid within days or weeks, at almost no cost.

Addgene's "zero-day, wetware-on-release model" thus played an essential role in shaping the uniquely meteoric uptake of CRISPR by academic biologists, magnifying the impact of the biological ease of construction by providing yet further simplification to construction thanks to a pre-existing distribution infrastructure that was able to cope with a high volume of demand. This structure ensured quickly turnaround of plasmids to requesters at low cost by obviating the need for experimenters to design and construct their own plasmids using internal resources or external contractors. Combined, these advantages reduced the burden of work required to get the new technology up and running in a lab, enhanced reproducibility of CRISPR experiments between labs by providing access to shared wetware resources, and generally enhanced the reliability of, and confidence in, CRISPR techniques amongst early adopters, who trusted the depositor's papers and Addgene's model and materials. This is not to say that Addgene's service is the only one available - numerous companies now offer more expensive and complete CRISPR kits which reduce the burden of work further by providing already-loaded GEVs containing custom guide sequences - but the availability of this service certainly made the notion of "trying out CRISPR" more appealing to academics, a population willing to publish their results, share additional materials, and drive further interest by the field at large.

It is noteworthy that innovators in ZFN and TALEN technology pioneered this distribution-consortium approach (Maeder et al. 2008; Maeder et al. 2009) to provide a roadmap for future CRISPR distribution efforts. Genome engineering pioneer Dr. Keith Joung noted that *"when we started the Zinc Finger Consortium in 2005, a major goal... was to make the technology available to all academics.... Addgene enabled us to do this [with CRISPR] efficiently and effectively [and] I believe that the tone and practice we established early on in the field then led to others following suit as the TALEN and CRISPR/Cas9 technologies emerged"* (Joung et al. 2015). According to Addgene Executive Director, Dr Joanne Kamens, it is likely that *"no single lab could handle the tsunami of requests for CRISPR/Cas reagents"* (Joung et al. 2015). Indeed, numerous biology publications cite Addgene as the provider of their CRISPR plasmids (Thompson and Zyontz, 2017), and internal data provided by Addgene indicates that at least 112,000 CRISPR plasmids were distributed to experimenters between 2012 and the end of

2017, which has come alongside a reduced distribution of ZFN and TALEN plasmids, as shown in Figure G.<sup>3</sup>

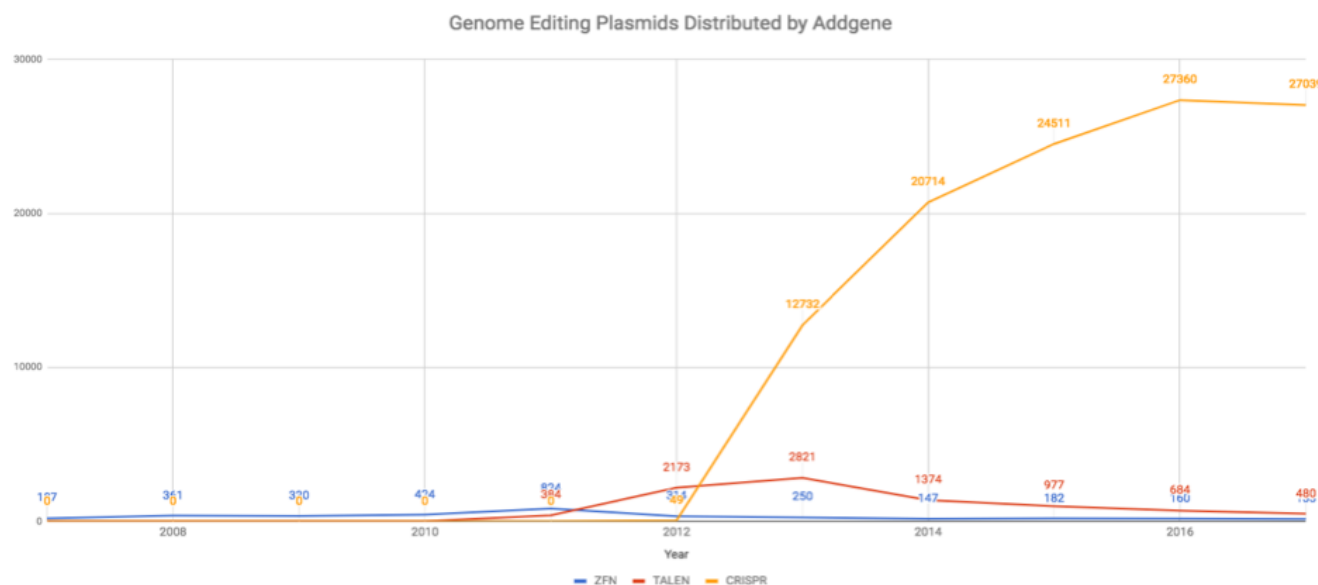


Figure G - graph showing annual distribution figures for ZFN, TALEN and CRISPR plasmids by Addgene between 2008 and 2017. Data courtesy of Addgene.

## The Virtuous Cycle

CRISPR RGENS have made it easy for researchers to edit genomes with lower investments of capital, per-experiment labor, and time costs in a broader range of organisms than allowed by older generations of genome editing tool. By shifting the foundational portion of the genome editing process from a multi-step protein engineering project to a one-step cloning project, the biological and laboratory barriers were lowered. These factors alone provided a step-change in capability, elevating CRISPR to the position of genome editing tool of choice - in a sense the tool de jour that any biologist would want established in their laboratory. However, CRISPR's outstanding success has been reinforced by the permissive intellectual property environment, and low-cost distribution infrastructure available to academics via Addgene. These factors were instrumental in ensuring that the early-adopter interest was met with sufficient supply, guaranteeing wide and rapid uptake. Together these factors have combined to promote high rates of publication of CRISPR papers, stoking further interest amongst potential users. Throughout the scientific and general media, the explosive CRISPR uptake has been referred to as a "CRISPR craze," but such notions fail to recognize the self-sustaining and positive feedback dynamics of this phenomenon. The dominance of CRISPR in the gene editing field is perhaps better described as a virtuous cycle (see Figure H) and has shown little sign of abating as the tool

<sup>3</sup> Plasmid distribution data used for Figure X analyses was provided by Addgene and is correct as of February 2018. Correspondence on file with author.

cements itself into the routines of many biotechnology laboratories and the minds of those that work in them.

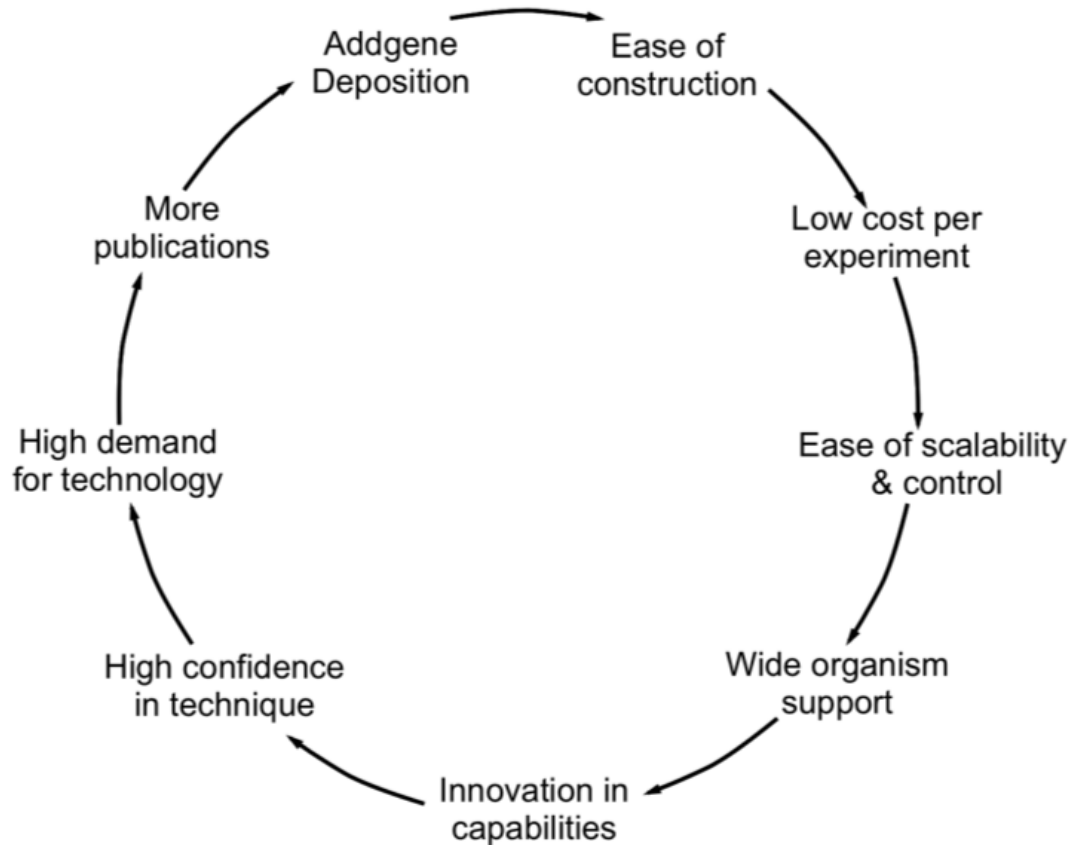


Figure H - schematic of the virtuous cycle underpinning the evolution and rapid adoption of CRISPR RGEN technology amongst life scientists.

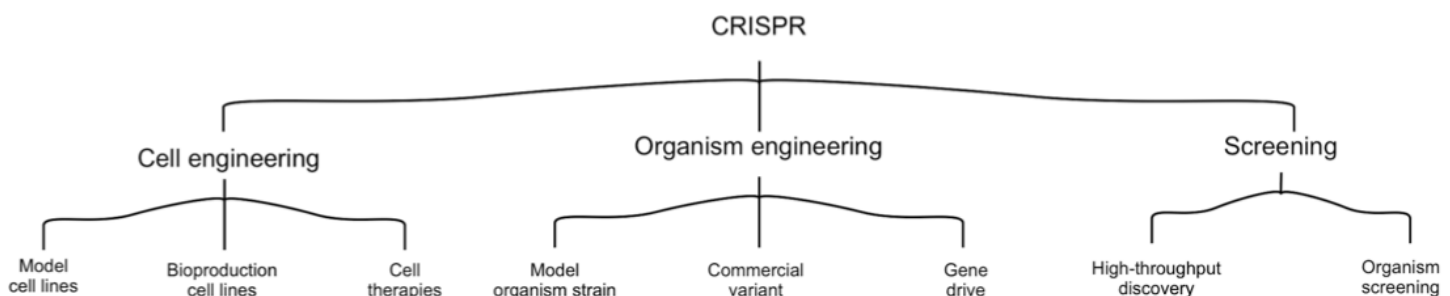
Whilst this perspective is bullish on CRISPR, it's important to note that CRISPR's adoption has not been universal amongst biologists. Even today "...researchers who believe CRISPR would be beneficial for their research might delay in adopting because the tool is still developing. Rather than being an early adopter today, such researchers could wait until CRISPR development matured, potentially making adoption easier and perhaps yielding a more powerful tool... Because of the large benefits of CRISPR, but also its continued development and shortcomings, researchers have reasons both for and against becoming an early adopter" (Thompson & Zyontz 2017). Moreover, as the applications of CRISPR advance beyond genome engineering, and into diagnostics and other tools, new branches of the CRISPR field will appear, with their own characteristics and trajectories that lie outside the scope of this analysis.

## SECTION 3: CRISPR - EMBODIMENT & PROCESSES

This section seeks to explain ways in which one can access and make use of CRISPR technology, describes the main variants of CRISPR genome editing applications, and describes an idealized version of a standard CRISPR process, documenting the steps that are conserved across the various families of CRISPR genome editing activity. Finally, the section explores the forms in which CRISPR technologies are embodied, be it software, hardware, wetware or something else.

### CRISPR, What Is It Good For?

There are three broad families of activity that CRISPR is used for: cell engineering, organism engineering, and screening. Each of these activities shares common applications across basic and translational research in academic, commercial, and biomedical settings. Cell and organism engineering have a variety of applications in these theatres, and the products of these processes tend to have direct value, whereas screening processes tend to be used primarily in a discovery setting, generating knowledge that can be indirectly exploited and applied by additional experiments and processes.



*Fig 1. A schematic of the basic taxonomy of CRISPR genome editing processes: cell engineering, organism engineering and screening.*

### Cell Engineering

Cell engineering is principally concerned with the deliberate and rational development of strains of cells so that they will have particular properties derived from particular mutations. Genome editing is used to introduce relevant mutations, enabling the development of ‘cell lines’ that can be used in research, the production of high value compounds, or as clinical cell therapies. A desirable quality of a cell line is that its genotype stays ‘stable’ over successive generations - in other words, as a parental cell divides, its daughter cells maintain the same genome and mutations as the parental lineage.

Cell engineering tends to be a highly deliberate activity, where investigators usually know exactly what mutation(s) they want to introduce. Cell engineering is usually an activity that exploits pre-existing knowledge to achieve a predictable outcome, although there may be a fair amount of experimentation and optimization with successive rounds of genome editing in order to ensure the behavior of the resulting cells is up to any task required.

With rational approaches, model cell line engineering supports the exploration and discovery of traits in basic and applied research, and is especially useful for deeply probing functional relationships between a mutation and a trait (Niu & Wang 2015). This is often referred to as mapping the relationship between genotype and phenotype. Cell line models have been used to determine likely clinical responses to drugs, and to understand how known genetic variations differentially respond to certain environmental conditions. Cell engineers might be more explorative in their work in cases where they lack guiding knowledge. Researchers might, for example, introduce redundant edits at different points of the same gene, to determine if knocking out a gene in different ways endows the cell with equivalent properties, helping to obtain a higher resolution understanding of how sequence information maps to a trait. More generally, model cell engineering is especially useful for subject matter experts, who adopt this process to provide meaningful answers to research hypotheses at the cutting edge of their chosen field.

Cell lines may also be developed as advanced bioproduction platforms, in which their metabolic pathways are altered in order to overproduce an already-present metabolite that is useful as industrial feedstock or pharmaceutical compound, or to insert novel production pathways for the same. Genome editing capabilities are used to enhance yield (Grav et al. 2017; Lo et al. 2017), either by modifying or introducing one or more genes to ensure the cell can synthesize the product, or by removing a cell's ability to shut off or alter the production once the compound reaches a certain concentration.

Cell therapies are medical interventions that inject intact, living cells into patients for clinical benefit. Broadly there are two ways an engineered cell can exert an effect on target tissues - recruitment of healthy cells, or removal of diseased tissue that is interfering with normal function. Typically, there are cells available in a living organism that already execute such desirable tasks, and genome editing allows these or other cells to execute those tasks in specific combinations, at specific sites. Cell engineering has thus unlocked a fairly novel and uniquely adaptive class of medicines that simultaneously detect and treat the underlying processes of a disease in ways that no other class of treatments has yet achieved. Notably these proposed therapies largely remain in preclinical or clinical development, and focus exclusively on somatic genome editing, such that "corrected genes" cannot be passed from a patient to their children. Depending on the territory in question, regulations on clinical CRISPR therapy tend to be highly restrictive, in-line with other medical interventions.

Prior to CRISPR, model cell line development was by-and-large conducted only when it was essential, and often at great cost of time or labor. Several commercial cell line development companies, and many academic *core facilities* provide cell line engineering capabilities to researchers as needed, especially in mammalian and human models. However, the advent of CRISPR has made model cell line development directly accessible to labs with fairly basic facilities and budgets. The same cannot be said for bioproduction or cell therapy - although the means to develop bioproduction or therapeutic cells are now widely available, and can be achieved in basic labs, bioproduction processes are best conducted in facilities that can support large scale production of valuable compounds in large bioreactors, which have significant operational overheads. While the quantities of cells required to develop cell therapies are minute, and early research may also be conducted in a fairly basic laboratory, taking these therapeutic engineered cells to market also requires dedicated facilities and approval from national or international regulatory bodies.



## Cell Engineering

	Model Cell Line Engineering	Bioproduction Cell Line Engineering	Cell Therapy Engineering
Description	Create a stable mutant cell line with traits and modifications that can be further studied to understand a state, disease etc.	Create a stable cell that produces high value chemicals, feedstocks, pharmaceutical compounds etc.	Create a living package of medicines that can adapt to a disease to exert therapeutic effects in patients
Classic Value Proposition	Conduct early stage investigations of physiology, molecular biology, and pathology, without the need to use higher organisms	Reduce the number of steps or intermediaries in a synthetic production process by building an organism that metabolically generates the desired compound	Directly treat a disease using modified cells from a patient, or standardized "off-the-shelf" cell therapies
Facilities Required	Can be produced in experimentally-relevant quantities in standard lab setup, scales well with facility size	Can be produced in experimentally-relevant quantities in standard lab setup, requires dedicated bioprocess engineering and bioproduction facilities for industrial scale extraction of compounds	Can be produced in experimentally-relevant quantities in standard lab setup, requires dedicated facilities and approved processes for clinical-grade production
Somatic/Germline	Typically somatic, may be used for germline	Typically somatic	Typically somatic
Service Availability	Services available from academic and commercial providers	Services available from academic and commercial providers	Largely developed in-house, some commercial services available

Figure J - an overview of common applications of the CRISPR cell engineering process.

## Organism Engineering

Whereas cell line engineering is used to produce a population of cells that is a whole product, whole organism engineering is concerned with generating living organisms with particular qualities. In many cases, a key objective is that the generated organisms are able to reproduce amongst themselves to produce a stable lineage of offspring with those same qualities maintained over generations. To achieve this, breeding programmes may occur over multiple generations in conjunction with genome editing. Whole organism engineering processes are typically executed by directly injecting (or attaching) germline (sperm, eggs and embryo) cells

with CRISPR payloads, which give rise to a whole organism at the end of a conventional reproduction process, and this may be supported by in vitro fertilization (IVF) in animals. It is the modification of germline cells that ensures that the edited genes are heritable. However, in species where whole organisms can be clonally derived from somatic cells, for instance plants, somatic cells may be edited, and are still able to give rise to a stable adult cloned organism.

Whole edited organisms have value in research and are commonly used as models for basic and translational research. Once genotype-phenotype relationships are established in cells (considerably cheaper and easier than whole organism engineering), researchers might develop an animal model to determine if the results are conserved in the whole organism. This is particularly useful in understanding the potential non-obvious effects that editing one system may have on another system's performance. To this end, organism engineering is valuable for proving and disproving scientific hypotheses. Organism engineering also has high value in creating new animal strains that are engineered to have disease genotypes, allowing higher fidelity testing and evaluation of new drugs and vaccines than would be possible in cases where diseases or their symptoms are phenotypically-induced in animals.

Whole organism variant engineering is especially useful in agriculture and other sectors that derive value from living organisms. Crop and animal species alike may be modified to have particular fitness traits improved, or otherwise modified to facilitate their productive husbandry and yield. Such translational and commercial efforts are reliant on the exploitation of pre-existing knowledge that is derived from basic scientific enterprise in universities, companies and elsewhere. Variant engineering can be practically considered as equivalent to the development of specific strains, breeds or cultivars, and CRISPR can significantly reduce the time and cost burdens associated with more traditional approaches. It is noteworthy that given sufficient time and budgets, the impact of the off-target issue can be effectively worked around in whole organism engineering. For instance, in the case where on-targets and off-targets are identified in an engineered organism, these individuals can be bred with variants that are not edited in order to obtain a portion of cross-breeds that have the on-target edit, and no off-target edits. For more details on the agricultural applications of genome editing, please refer to the issue brief of this volume authored by Dr Sarah Carter.

Whole organism variant engineering is not confined to farmed species - pests and targeted wild species may also be the subject of genome editing, either for experimental research or for translational purposes. CRISPR has, in particular, advanced the development of, and discussion around, novel applications for gene drives, a technique in which an initial lab-developed individual organism may be released into the wild to drive the inheritance of a particular trait across an entire population, discussed in more detail in Prof. Kevin Esvelt's contribution to this volume. Whole organism genome editing is being evaluated outside industrial settings for its potential use in conservation, either as a tool to assist the understanding of organisms' responses to conservation challenges, or as a potential method to advance an organism's resistance to environmental pressure or modify a small population's vigour by enhancing genetic diversities (Novak and Maloney 2018).

The creation of whole organisms using genome engineering is likely to be highly regulated and subject to strict controls. The nature of this regulation depends on the organism and country but animals and plants that are intended for human consumption are typically the target of the most regulation, with decorative organisms and pets subject to reduced regulation.

At present a suite of policy proposals and regulations regarding CRISPR gene drive organisms and conservation applications are emerging from academic and government bodies alike.

Finally, it is notable that whole organism engineering can also apply to humans, especially in a medical context. Since many countries have declared a moratorium on this work or already have pre-existing laws in place and no commercial market for this type of genome editing exists, it has an unknown future.

## Organism Engineering

	Model Organism Strain Engineering	Commercial Variant Engineering	Gene Drive Engineering
Description	Create a reproductively-competent whole organism for further study	Create a reproductively-competent whole organism for commercial production of crops, meat etc.	Create a reproductively-competent whole organism for ecological intervention
Classic Value Proposition	Conduct mid-late stage investigations of physiology, molecular biology, and pathology, in an organism specially bred to mirror a human disease	Improve the yield of an organism currently used as the source of agricultural, commercial, or industrial products, or modify an organism to compete with naturally-occurring variations that may cause disease	Potentially a highly effective means of spreading a specific trait through a population over many generations, without the need to supplant the existing population - more 'hands off' than traditional release of modified organisms
Facilities Required	Can be produced in experimentally-relevant quantities in standard lab setup with appropriate organism culture capabilities; scales well with facility size	Can be produced in experimentally-relevant quantities in standard lab setup with appropriate organism culture capabilities; scales well with facility size	Can be produced in experimentally-relevant quantities in standard lab setup with appropriate organism culture capabilities; scales well with facility size
Somatic/Germline	Typically germline (some organisms can be cloned from somatic tissue)	Typically germline (some organisms can be cloned from somatic tissue)	Typically germline (some organisms can be cloned from somatic tissue)
Service Availability	Services available from academic and commercial providers	Services available from academic and commercial providers	Not specifically available from academic or commercial providers

*Figure K: an overview of common applications of the CRISPR organism engineering process.*

## Screening

High-throughput screening (HTS) processes make use of CRISPR in conjunction with heavily automated liquid handling platforms, measurement devices, and cell culture capabilities to conduct many millions of experiments simultaneously. In many ways the basic cell biology occurring in these processes is the same as what has been described above, the primary difference is the scale and scope. Whereas cell/organism engineering experiments tend to be highly focused, rationally exploiting existing knowledge to obtain a specific genetic variant or lineage, screening experiments often modify large populations of cells at multiple genomic loci in order to discover complex, multi-gene phenotype-genotype relationships for later exploitation. A helpful general characterisation is that HTS is largely discovery-oriented, making use of automation to explore a wider problem space than achievable with more manual cell engineering activities, which tend to deploy knowledge derived from discovery processes in order to engineer a specific cell line or organism.

Cells are exposed to a *genome editing library* - a scalable collection of GEVs - designed to interrogate the function of many target genes or regulatory genomic elements. In practice, a portion of the cells receive one or more CRISPR guide RNAs from the library, which binds to a corresponding target sequence to effect a change in the genome. Cells are typically exposed to a selective pressure (a drug or an environmental condition) that may cause edited cells to survive or die at higher rates than unedited cells. Different combinations of mutants are generated across the entire population, and the cells are then evaluated with next generation sequencing (NGS), automated cell sorting, and/or other methods to connect sequence-level alterations with observed phenotype. Sequencing allows researchers to see which cells dropout or expand in population size as these stimuli are applied. To elucidate these genotype-phenotype relationships with a CRISPR library, it is critical to ensure that every guide in the library is likely to successfully interact with the target site to decrease noise and identify true positive events. In this context, off-target edits can be an issue since they generate a significant amount of noise in an already-complex experiment. Engineered cell lines with specific traits may be subject to a screen in this way, or alternatively cells with interesting properties (e.g., primary cancer cells derived from a patient) may be used. Single celled organisms might also be used.

Rapid low-cost construction of GEVs is especially useful in the context of HTS. Indeed, screens with the scope, accuracy, and affordability offered by CRISPR were not really possible with precursor gene editing technologies. Alternative RNAi technology, a gene silencing method that does not make permanent changes to a genome, but has relevance to screening, has considerable off-target issues (Jackson et al. 2003; Jackson et al. 2006; Jackson & Linsley 2010), putting CRISPR in prime position to dominate this process.

Screening experiments may be performed in pooled or arrayed formats. In a pooled format, a population of cells is exposed to the entire library of guides at the same time, in the same tube (notably, sub-pools may also be used). In an arrayed screen, several distinct populations of cells are exposed to distinct guides from the library, with no population receiving more than one guide, with the experiment commonly executed in one or more multi-well plates. Pooled screens are useful for discovering traits for the purposes of functional genomics and target identification, whereas arrayed screens are typically used for follow-up work and target validation in drug discovery, ahead of cell line or model organism engineering as a final validation. HTS approaches are best used in microbial or cell-based projects, where the required

volume of genetically engineered cells can be kept within a small facility housing many dozens to thousands of multi-well plates. While these cells could potentially be grown into whole organisms (if the species lends itself to this), the practice is rare as this could quickly result in millions of large organisms outgrowing a facility. As such, a screening funnel is usually applied, with pooled experiments on many cells preceding arrayed experiments on fewer cells, followed by whole organism engineering of the smallest subset of interesting genetic variants.

At their core, HTS activities have functional genomics in mind, and overlap with low-throughput cell engineering activities. The intent of separating these subjects is to emphasize how important scale is to HTS as well as its unique requirements. HTS is highly specialized and expensive, and is typically out of reach for basic or mid-sized laboratory facilities. The activity of discovery is of great importance in the commercial biotechnology, pharmaceutical and agricultural sectors, which can marshal the resources to execute all aspects of this process, and must do so in order to compete with their peers in the commercial discovery space. While the advent of CRISPR has certainly reduced the biological limitations of this process by making larger libraries easier and cheaper to assemble, the operational requirement to invest in the automated facility still puts this out of reach of many actors, a subject discussed further in Section 4.

## Screening

	High Throughput Discovery	Organism Screening
Description	Systematically modify the genomes of cells to identify genomic loci that code for specific traits	Systematically modify the genomes of whole organisms to identify genomic loci that code for specific traits
Classic Value Proposition	Conduct very large-scale investigations of physiology, molecular biology, and pathology, without the need to use higher organisms	Conduct medium-scale investigations of physiology, molecular biology, and pathology, in whole organisms
Facilities Required	Can be executed in a high-throughput molecular biology lab setup that is specialised to this task. Requires significant efforts in sequencing and data processing	Can be executed only at facilities dedicated to culturing organisms at large scales, e.g. dedicated agricultural biotechnology farms. Benefits heavily from automation
Somatic/Germline	Typically somatic	Typically germline
Service Availability	Services available from commercial research providers	Largely developed in-house, some commercial services available

*Figure L - an overview of common applications of the CRISPR screening process.*

## Other CRISPR Applications

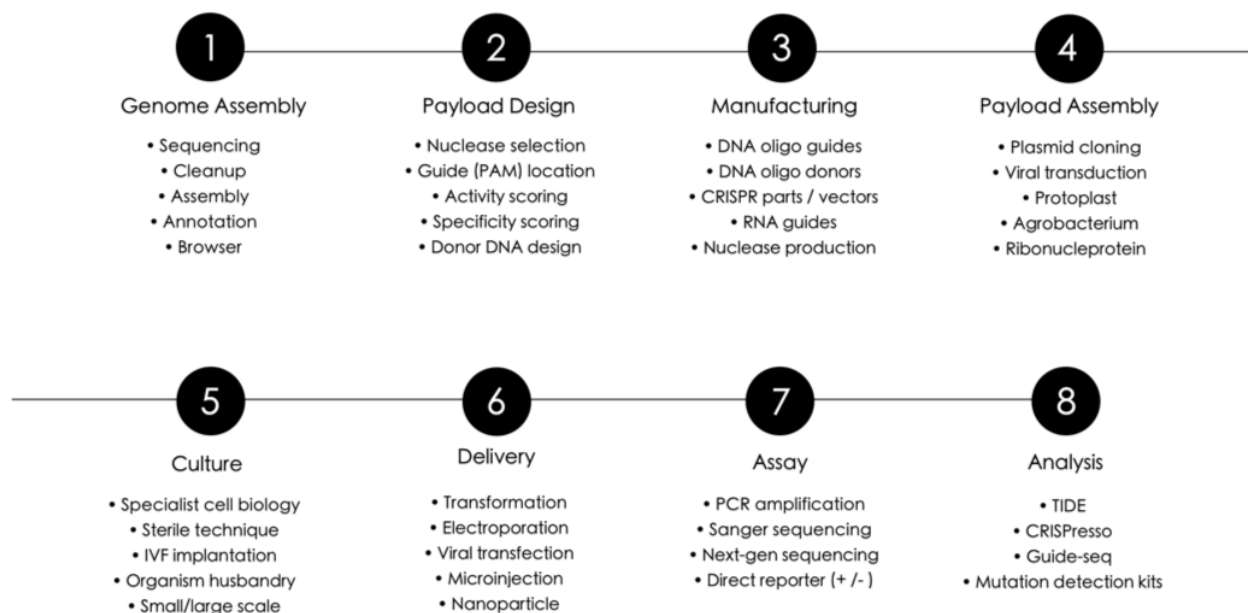
The three families of activities discussed above are illustrative of the better-known contemporary applications of CRISPR genome editing technology, but they are not exhaustive. CRISPR is increasingly being used beyond the scope of *editing* and applied to other interesting problems in biology, such as biosensing and diagnostics. CRISPR approaches have been used to detect picomolar amounts of RNA without the need to amplify material in a lab (East-Seletsky et al. 2016), but have also been combined with thermal amplification techniques to achieve single molecule detection (Zuo et al. 2017). The SHERLOCK system (Gootenberg et al. 2017) has been shown to detect ~1000 viral particles in 1 ml of solution, and ~2000 copies of Zika virus in clinical isolates, with high specificity and single-base discrimination (i.e. it is able to discriminate between the African and American strains of Zika virus). CRISPR systems appear to be well-suited for diagnostic and detection applications since their components are suitable for freeze-drying and rehydration without sacrificing activity, obviating the need to rely on cold chains for storage and shipment. Please refer to the issue brief by Dr Kyle Watters, in this volume, for further details on the global health security applications of genome editing.

## Process and Embodiment

### Overview of the Ideal CRISPR Process

CRISPR genome editing is not simply about the act of cutting and repairing the target DNA, but the events leading up to, and beyond, those moments, including predictive design activities of programmable elements, and measuring the outcome of the work. Genome editing is thus not a discrete activity, but rather a generalizable process, with many steps and possible combinations of steps, all depending on the applications and desired outcome.

An archetype or ideal CRISPR genome editing process, as shown in Figure M, is a helpful extension of concepts, outlined in Figure A, and is intended to be useful for considering processes for future programmable nuclease tools, as well as CRISPR itself.



*Fig M. The ideal CRISPR process with each element representing a distinct step in an average genome editing experiment, and the possible techniques or decisions that a researcher might consider when executing that step.*

A complete end-to-end execution of the ideal CRISPR process requires a wide range of skills, and more involved genome editing applications, such as screening, will be far more complex processes (with potentially many more steps) that are challenging to complete successfully. Equally, a more routine genome editing procedure, for instance a simple knockout of a gene in a well-studied organism, will likely be a correspondingly simpler process, with fewer steps.

CRISPR technology is embodied in many forms that extend beyond nucleases, guides, and GEV formats, also including the software, hardware, and commercially available kits and services which exist to facilitate these steps.

Depending on project complexity, either an individual or a team will take on one or more of these steps, each requiring some specialist training and access to one form of CRISPR technology or another to successfully complete the procedure. Each step will require a distinct set of laboratory tools and techniques, requiring the user or team to be comfortable using molecular tools and delivery techniques, maintaining cells or organisms over extended time frames, using various assays, bioinformatic design tools or analysis packages, and employing specific assays and commercial services. It is noteworthy that not all CRISPR users will care to execute each step to the same degree of accuracy, and that this may be due to a deficit of the required skills, facilities, or funds, or otherwise simply due to a user's decision to deem one or

more steps unnecessary for their particular use case. In other words, a user can execute the cleanest and most complete CRISPR genome editing process by fulfilling each step to the highest possible standard, or could settle for a “quick and dirty” approach that is good enough to work, but might not stand up to academic peer review, or achieve an outstanding outcome.

One should also note that the connection of steps within the entire genome editing process is not always a simple affair. Equipment and reagent manufacturers and other technology developers tend to create solutions for a particular step, or adjacent steps, rather than the entire process, and data-transfer, and biological compatibility problems may need to be troubleshoot. For instance, data may need to be converted from one form to another, interpreted by distinct specialists or outsourced to external service providers. Investigators may need to develop their own protocols to transfer samples between manufacturer protocols, and in many cases a protocol, reagent or kit may only be suitable for a particular organism, requiring interpretation and adaptation.

A good deal of complexity is encountered when attempting to describe each of the possible routes that one could take to obtain necessary components to perform a CRISPR experiment, as not only are there many dimensions to this problem, but also many service providers and kit producers who offer CRISPR products in different forms that bridge these steps, or otherwise make them irrelevant from the user’s perspective.

## **Detail of Technology Embodiment for Each Step of the CRISPR Genome Editing Process**

### **Genome Assembly**

To edit a genome with meaningful accuracy, a representation of its sequence information must be available as a digital model, known as a reference genome. A good deal of model organisms and economically-relevant species have had their genomes sequenced and made available online, and researchers can directly access this data to use for computational design of CRISPR payloads. CRISPR users can also access genome data through dedicated genome browser tools that represent genome data visually. In some cases, the genome of more esoteric organisms may not be widely available, or only partially sequenced, and researchers will need to conduct a more involved genome sequencing project. This will involve obtaining the organism, extracting DNA from its cells, sequencing the DNA with specialist tools, and deploying bioinformatics expertise to assemble the different reads into a contiguous whole and annotating (sorting and labelling) the salient genetic sequences to differentiate between different genes. This process is most intensive when working with an organism with no pre-existing genomic sequences available, but made easier with access to pre-existing assembly and annotation knowledge about the organism, or a closely related species. In some cases, where a reference genome is not sufficient for confidence, notably in clinical applications of CRISPR where editing activities must be precise, researchers may choose to re-sequence the genome of the exact cell or organism that they are working with, in order to obtain a digital representation of their chosen model, rather than trusting pre-existing data obtained from different individuals of the same organism, which will not accurately represent the soon-to-be-edited cells.

A number of providers offer whole genome sequencing and assembly services. *De novo* sequencing is expensive, and scales with the size of the genome and the complexity of the



project. Re-sequencing is much cheaper as it builds off an already-established reference dataset. A number of universities offer researchers access to sequencing core facilities that are able to provide in-house genome sequencing, assembly and annotation services to affiliated labs at low costs, and unaffiliated labs at higher costs. Finally, a number of bioinformatics genome assembly and annotation toolkits are available via open and closed source distribution models, allowing researchers to input the gathered data from sequencing devices into their own bioinformatics pipelines for assembly and annotation. Finally, a number of online and offline genome browser software tools are available, which provide researchers with convenient digital representations of reference genome datasets, and typically provide capabilities to upload arbitrary genomic datasets for visually inspecting genomes at different scales (genes, chromosomes, genomes).

## Design

With sequence information in hand, the genomic targets of interest can be identified, a general genome editing strategy established, and the elements of a CRISPR payload designed. A CRISPR RGEN will be selected based on the researcher's intended goal and the predicted quality of nuclease-genome interactions. Key to this choice will be the availability of PAM sites (NGG sequence motifs for SpCas9) close to the intended cut site, the lack of which will prevent the use of that nuclease, and require the use of another. Once a target cut site is identified, the final set of guides can be decided upon, usually three to five, which ideally have high performance off-target and on-target activity scores. Scores are assigned to each guide based on its sequence, using bioinformatics algorithms trained on the performance of similar sequences in historic CRISPR experiments used to calibrate the scoring algorithm. In general bioinformatic CRISPR design tools allow experimenters to stand on the shoulders of calibrations work conducted elsewhere, and a great variety of free-to-use scoring functions are available, some requiring a user to download and install source code and be comfortable using a command line interface, others available online with varying levels of attention paid to user interfaces which facilitate guide selection (Doench et al. 2016; Stemmer et al. 2015; Montague et al. 2014; Hough et al. 2017; McKenna & Shendure 2018). Researchers will typically select these design tools based on the number of citations for that tool, and peer input from their lab. Not all CRISPR users will use all of the design tools, but it is common for multiple tools to be used to measure different attributes for the same experiment in order to identify the top-performing guides across several scoring functions. CRISPR users who are inserting exogenous DNA in a knock-in experiment will also design appropriate donor DNA, typically using a standard DNA computer-aided design (CAD) package, or a CRISPR software tool that has this feature. Some CRISPR payload design tools also provide integrated genome browser capabilities, blurring the line between the genome assembly and design steps, however these design tools tend to have very limited support for different genomes, especially non-model organisms or those of limited commercial value.

Design considerations are difficult to separate entirely from assembly and delivery, as the CRISPR payload and GEV will need to be designed with these strategies in mind. The successful completion of this step makes ready a complete set of digital designs for the CRISPR GEV and guides to be used in the experiment.

A number of commercial providers offer CRISPR payload design services, either as distinct services, or as part of a package that includes DNA manufacturing and CRISPR payload assembly, with the design tool itself incentivising a user to purchase reagents through a one-stop-

shop. The vast majority of CRISPR users will make use of existing design tools published by leading academic labs, but some advanced users with appropriate bioinformatics capabilities may seek to develop their own design tools for specialized use cases (related to a particular organism or RGEN variant).

## Nucleic Acid Manufacturing

Designs can be easily represented as digital sequence information, which can be sent to DNA synthesis companies for chemical production of the relevant guides and other coding DNA sequences, often at very low cost and in short time frames for simple experiments. The requirement for manufacturing depends heavily on what the user already has available in house, the nature of the delivery approach they intend to take, and the degree to which they might attempt to source pre-existing constructs. For instance, a user who already has access to a CRISPR expression plasmid sourced from Addgene, will only need to manufacture the guides in DNA form. If they intend to modify that plasmid, they may need to have other DNA elements of the payload synthesized ahead of an assembly step. On the other hand, someone using a RNA-based GEV will not need a plasmid or DNA synthesis per se, instead requiring an RNA synthesis service, or an RNA expression kit. Nucleic acid manufacturing is rarely handled by CRISPR users in-house as a number of service providers already compete on price to serve the global life science market. Nonetheless, certain large companies or central facilities at academic centers may opt to handle their nucleic acid manufacturing with in-house synthesis hardware if the economics make sense.

Nucleic acids may be sourced from DNA synthesis providers in the form of unmodified DNA, or chemically modified DNA or RNA, arriving in the form of individual tubes or in a plate containing wells of nucleic acids. Commercial providers may also distribute nucleic acids in kit forms, which contain other CRISPR reagents, plasmids etc., needed to ensure GEVs are appropriately assembled or otherwise made experiment-ready. Typically, kits will command higher prices as they provide greater value to users by simplifying multiple steps in the process. Users may obtain limited batches of nucleic acids for small projects, or obtain large batches in the form of libraries, either specifying their own designs, or purchasing pre-designed or pre-manufactured libraries of many hundreds or thousands of guides.

## Assembly

The assembly step constitutes an edit-competent CRISPR construct from a GEV, nucleic acid and any other payload or packaging components, ready to be delivered to the target cell. The complexity of assembling the final CRISPR construct (or constructs) depends entirely on the nature of the delivery approach that is taken and the approaches taken in previous steps. Guide nucleic acids can be fairly easily cloned into a plasmid using very basic molecular biology skills and simple protocols. Assembly of RNA requires little work as RNA is by-and-large deliverable as soon as it is produced. Assembly of CRISPR-expressing viruses however may be more difficult as this typically requires a better-rated laboratory that can handle the additional steps required to transform bioproduction cell lines into virus factories, and then also to harvest the virus and concentrate it ahead of transduction (delivery of viruses to the target cells). Viral work can add several days or weeks to a project. As for synthetic RNP, assembly is again a fairly simple operation in which Cas proteins and pre-transcribed RNAs are complexed *in vitro* ahead of the delivery to the cell. In some cases, CRISPR users may optionally choose to verify that

their assembly was successful with a quality assurance sequencing run of the assembled DNA constructs, with follow-up analysis to provide insight on the quality of assembled GEVs.

Assembly may be optional or unnecessary if a significant portion of the project is outsourced. For instance, a user intending to use a virus-based delivery approach may opt to manufacture and assemble the viral GEVs themselves or may employ a service to handle both steps simultaneously. Likewise, guides and plasmids may be synthesized and cloned by a specialist provider - or even designed by one. When thinking about manufacturing and assembly it is helpful to keep in mind that the raw complexity of the steps tends not to be the issue -rather it is about managing the scale of work required by the team available, which becomes especially important in large-scale screening-style experiments involving hundreds of thousands of constructs. As mentioned above, and described in more detail below, a large number of pre-designed, manufactured, and assembled CRISPR libraries are available to meet the needs of HTS users.

## **Culture**

Cells and whole organisms have specific husbandry and culture requirements that must be adequately addressed throughout the genome editing process. Nutritional media and other life support requirements like thermal control or oxygen provision, must be met to maintain populations of cells or microbes in dishes or bioreactors or live organisms in vivariums or enclosures. Cell biology requires sterile technique to be observed to ensure cultures do not become infected. For small scale genome editing work with cells, much of this can be achieved with what is available in a basic laboratory setup, but will become increasingly complex with scale, requiring additional staff or investment in automation. Whole organism culture will depend entirely on the organism being engineered - maintaining gene edited fish will obviously have different requirements from maintaining gene edited plants or insects, which also have different requirements in terms of breeding programmes which might involve protracted rounds of IVF and veterinary expertise in higher animals. Successful husbandry of organisms in lab conditions over extended periods is a major challenge requiring significant capital expenditure on facilities and expert staff before, during, and after the delivery of CRISPR constructs to the target cells. As described elsewhere, genome editing programmes may require multiple generations of organism to be generated and bred in captivity, and if no husbandry protocol exists, substantial time may need to be allocated to this endeavour.

Where individual labs are unable to meet their culture needs in-house, dedicated facilities are available in the form of laboratories and core facilities in commercial research campuses and universities, as well in the form of contract research organizations equipped to handle particular organisms or large scale bioproduction operations. However, it should be noted that the majority of early stage genome editing work will not require dedicated facilities, which will only become important with scaleup.

## **Delivery**

The approach to deliver the assembled CRISPR construct depends on the GEV used, as well as the organism, its life stage, and cell type (somatic or germline). Heat-shock transformation is a relatively simple affair in which cells are incubated with plasmid GEVs under alternating thermal conditions using ice and a water bath, allowing pores in cell membranes to

open up and allow DNA to enter. Electroporation achieves the same cellular outcomes, at higher efficiencies, by passing an electric current through the cells using an electroporator. Electroporation and heat shock transformation are useful techniques when working with a large number of cells and the percentage of successfully transformed cells does not need to be high. However, some cell types resist these transformation methods, and may require the use of viral approaches, in which cells are incubated with engineered viruses containing CRISPR constructs. Viral transduction is commonly used in mammalian cells, with commonly used viruses including lentivirus which randomly integrates the CRISPR construct into a cell line, and adeno-associated virus (AAV) which affords targeted integration. In plants, agrobacterium may be used as vectors to deliver CRISPR systems into target cells.

In cases where the target cells are limited in number, for instance when working with germline cells, microinjection may be used instead, which requires the use of a specialized apparatus to directly inject the cells with CRISPR constructs. This procedure can be automated at considerable expense. Once delivered, the CRISPR construct will boot up inside the cell, and begin introducing DSBs at target sites, stimulating DNA repair processes as intended. Some organisms, especially those with multiple cell walls and membranes, may resist these delivery approaches entirely, making them intractable targets for genome editing.

CRISPR delivery systems in all their forms are widely available from reagents providers in the form of individual, small batch products, pre-made kits that also deal with design and manufacturing steps, and custom services that will handle more complex delivery methods such as large-scale microinjection or viral transduction.

## Assay

The assay step seeks to obtain some form of readable output, in qualitative or quantitative form, to confirm the success of the genome editing procedure. Basic qualitative assays may use a direct reporter approach, in which an obvious change in edited cells' phenotypes will help CRISPR users differentiate between those that are edited or unedited, for instance a change in phosphorescence. Reporter assays may be positive or negative, in that they measure the presence of something new, or the absence of something that was present before. Alternatively, qualitative approaches may use a polymerase chain reaction (PCR) approach known as a restriction fragment length polymorphism (RFLP) assay, where DNA from the cells is amplified via PCR, and then cut using one or more restriction enzymes. The treated samples are visualized using an electric current running through an agarose gel, which separates DNA fragments of different lengths, with genome edited cells producing a different visual fingerprint from unedited cells that were in treated the same way. This allows researchers to identify and isolate different populations of cells for follow up work. Both of these approaches are simple to use, cheap, and available to those using typical lab setups. Qualitative reporter assays can also be automated, for instance with the use of flow cytometry and cell sorting equipment to physically separate edited cells from unedited cells in a high-throughput manner.

Quantitative approaches may use Sanger DNA sequencing methods to amplify targeted genome regions, usually on-target sites, and some of the off-target sites perceived as being higher risk to experimental outcomes. Redundant Sanger sequencing of select target sites is often enough to draw acceptable conclusions for early stage projects done with a small budget. Next-generation sequencing (NGS) approaches, which provide greater confidence in the results but are

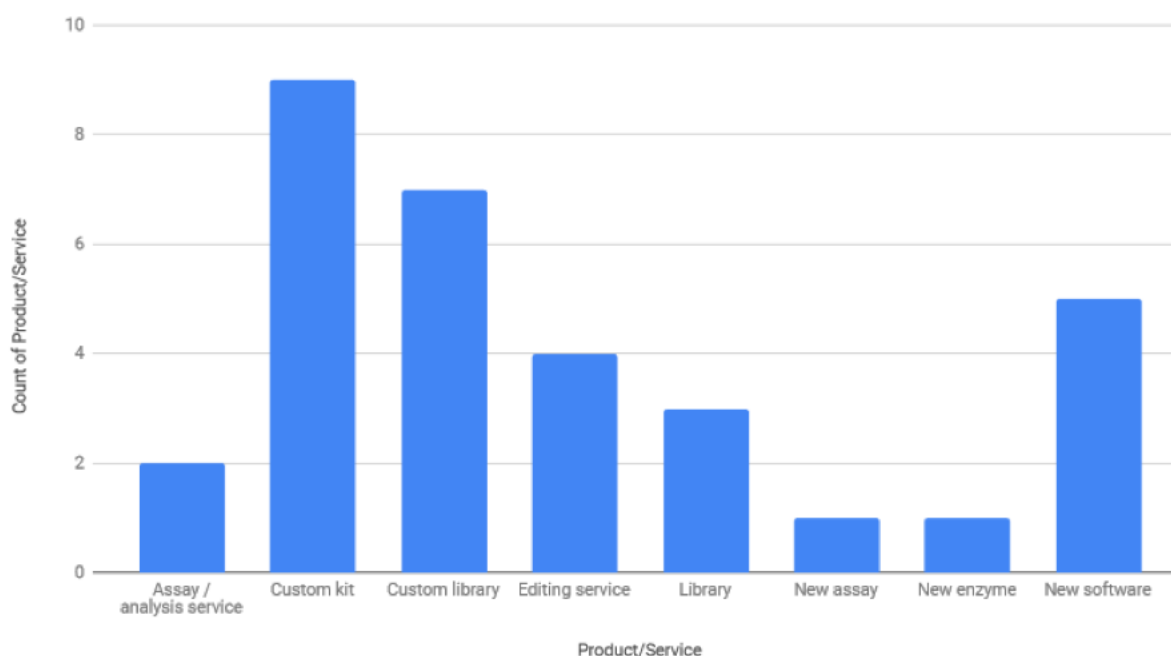
more expensive, are commonly used in translational and pre-clinical genome editing, although are increasingly found in basic research laboratories and projects. NGS approaches may be targeted to specific on-target or off-target sites, or alternatively may be used in the context of whole genomes, which provide the greatest confidence in editing outcomes. CRISPR users may use more than one assay technique, using cheaper and simpler techniques that produce qualitative data in the early stages of a project, and progressing to more expensive and challenging quantitative assay techniques in the later stages of a project.

Assay reagents are available in the form of small scale products that support in-house qualitative approaches, such as the standardized gels and restriction enzymes required for RFLP assays, or services for custom nucleic acid (primer) synthesis used with this approach, as well as for Sanger and Next Generation Sequencing. An increasing number of pre-made or partially customized kits are available for detecting editing activity, where users can specify the targets they intend to edit or avoid editing, and receive an all-in-one package. Demand for quantitative NGS approaches is met by a large number of service-providers and core facilities operating sequencing capabilities.

## Analysis

For many researchers, the results of an assay may be sufficient to conclude an experiment. In cases where a large amount of data is generated, for instance in pooled CRISPR screens measured via NGS sequencing, additional analysis steps may be required to validate findings and provide confidence in experimental conclusions or clinical safety. Bioinformatics pipelines are essential for analyzing complex data sets, serving as multi-part computer programs that take input data from an assay readout, transform it into a format that another program element can read and process, and pass onto the next computational worker in the chain. Bioinformatics pipelines effectively automate the complex data clean up and analysis operations required to obtain human-readable measurements from molecular biology experiments. They can be extremely complex and require specialist knowledge to develop and use effectively. The vast majority of CRISPR users will make use of an ever-growing list of publicly available genome editing analysis programs developed by academic labs, but some labs and many companies working on more esoteric genome editing experiments will develop their own pipelines for their own kinds of CRISPR experiments. These tools have different skill requirements. MAGeCK is a command line tool that allows users to identify metabolically important (essential) genes from CRISPR screens (Li et al. 2014). TIDE allows users to upload data from Sanger sequencing runs to quantify editing accuracy and the predominant types of mutations induced by CRISPR in knockout experiments (Brinkman et al. 2014), and a more recent TIDER tool can quantify knock-in experimental results (Brinkman et al. 2018). Like other bioinformatics tools, some require familiarity with a command line interface and scripting, whilst others can be more easily used with a graphical user interface. A number of analysis services are also made available by companies which offer free-to-use programs online, or otherwise provide this capability as part of a genome editing service.

CRISPR Product / Services 2012 - 2017



*Fig N. CRISPR products and service announcements issued via press release between June 2012 - January 2018. 142 press releases pertaining to CRISPR were analyzed from June 2012 - January 2018. 34 pertained to new products or services, shown here to illustrate trends in product and service development. Note: the data should not be interpreted as conclusive regarding the relative popularity of different products / services as it is drawn entirely from press releases, which bias organisations that are incentivized to be loud about a launch. As such the data does not indicate if these products were welcomed by CRISPR users beyond the initial launch, and we call for a more in-depth analysis in follow-up research.*

## CRISPR Kits and Other Commercial Trends in Embodiment and Accessibility of CRISPR Technology

Commercial providers have rapidly moved to colonize the CRISPR reagents and services market, offering numerous solutions for different steps of the CRISPR process. Whilst some laboratories will brew some of their own CRISPR reagents in-house, almost every user will need to source some elements from an external provider, namely nucleic acids. By analyzing 142 press releases related to CRISPR technology, products and services between June 2012 and January 2018, a striking picture emerges that indicates a proliferation of custom kits, custom libraries and novel software tools to facilitate the genome editing process.

The value proposition for a custom kit is that the user will receive a set of experiment-ready reagents, in the format of their choice, enabling genome editing out of the box. Typically focused on a handful of targets, the kits are partially custom in that users may specify a gene they want to target in an organism, or request a guide sequence directly. The manufacturer then converts the request into a nucleic acid sequence and synthesizes what's required, potentially also assembling this into a GEV or other functional CRISPR payload - all at a fairly low cost. A

variety of custom kits exist, bridging not only the design, manufacturing and assembly stages, as is the case for custom RNP kits offered by Synthego, but also even assay steps, such as the EnGen mutation detection kits available from New England Biolabs. By no means should these examples be considered exhaustive, and it should be noted that experienced users will typically be able to achieve the results promised by a kit without having to use one, but may still elect to use a kit to reduce their time commitment to a project.

Kits reduce the labor required to execute a particular step, potentially obviating the need for a user to have relevant skills in this area. However, kits do not solve the challenges pertinent to the other steps in the process, and as such kits can only blackbox a limited portion of the entire process. Moreover, kits can only solve well-defined problems that are common enough to make that kit commercially viable for the provider. Finally, kits are not silver bullets for the unpredicted complexities of genome editing, which may arise in any experiment at any time, and were not anticipated by the provider or user. As such, even when kits are used to facilitate one or more steps, they may not be adequate for troubleshooting the failure of those steps, which may require the experiment to be repeated from the ground up.

The value proposition for custom library service is that design, manufacturing, assembly of many thousands of GEVs may be outsourced, providing significant reductions in labor on the user-side, and dispensing with the tedium of producing thousands of GEVs in-house. Library providers tend to pass on the benefits of scaled production operations to the user, resulting in a cheaper cost per GEV as library size increases, at a price point efficiency that outstrips what an average user would be able to achieve in-house (assuming they did not have access to automated approaches). Moreover, users are also able to specify how a library will be designed, either choosing a subset of a pre-existing genome-wide library (Shalem et al., 2014), or alternatively providing instructions on which scoring function and guide design tool to make use of for designing entirely novel libraries. User specification of guide design rules thus ensures that a custom library is designed using the latest know-how, rather than older-generations of guide design rules.

## **Embodiment in People and Educational Resources**

In addition to physical hardware/wetware, and intangible software, CRISPR technology is also embodied in the tacit knowledge of experienced operators and the training materials and educational resources offered to novices and experts alike.

The capability of an individual or team of genome editors will depend on their experience and knowledge about the CRISPR tool itself, as well as each of 8 steps outlined in our ideal process. In this sense, both the explicit and tacit knowledge about the process may be held across a community. Users may be experts about the theoretical concepts of their genome editing tool (for instance payload design as it relates to CRISPR), but may be inexperienced when it comes to the assembly and delivery of lentiviral GEVs - a distinctly non-CRISPR step. For each of the tools, and steps in the process, users may be novices, intermediates, or experts, and they may collectively have gaps which may or may not be addressed by commercial / outsourcing services, software tools or kits. Individuals or teams are also likely to have expertise in the model organism and subject area they are using as the target of their genome editing activity, for instance a mouse model or a specific genetic disease. In smaller labs, there may be one “CRISPR specialist” who conducts all aspects of the genome engineering process alone. In larger labs,

people may specialize and a manager may oversee the entire process. The siloing of skills is particularly common in the largest labs, especially commercial labs where individuals will handle the same step for different projects, in an assembly line process. In this sense, a team of users operating a genome editing tool in the context of a larger process is only able to execute their procedure as well as the weakest link in their communal process. For successful genome editing to occur, it is ideal to have the highest level of communal expertise across all of these areas, which is more likely to be present in an elite institution than a lone amateur.

Tacit knowledge is important to all disciplines of molecular biology, and experience is best gained through practice and mentorship by those experienced in the art. The wide availability of protocol manuscripts in different species, free step-by-step guide books offered as promotional content by commercial providers, conferences and webinars dedicated to different applications of genome editing, alongside numerous online fora, all provide further opportunities for users to exchange best practices or learn basic and advanced techniques. However, nothing is more informative than on-the-job practical training, especially when one has the opportunity to directly troubleshoot problems with guidance from a lab instructor.

A review of the literature suggests there have been no coordinated studies documenting the demographics of CRISPR or other genome editing tool expertise across or within laboratories, but globally the US leads the way in the number of CRISPR publications, followed closely by China, Japan, the UK and Germany (Siwo 2018). It is likely that these nations host the most well-established centers of excellence and knowledge-generation for the application of this tool, and would be ideal choices for CRISPR novices to relocate to in order to acquire tacit knowledge and experience in genome editing. The concept that “anyone can do CRISPR” or “CRISPR is easy”, with no experience with other molecular biology techniques required for the process, is likely a fiction: a non-expert journalist attempted CRISPR and failed at a basic laboratory step (Cohen 2016). Regardless, undergraduate students are able to learn everything needed to successfully use CRISPR to generate mutant fruit flies within 15 weeks (Adame et al. 2016), illustrating just how low the skills barriers to entry are to achieve basic results with this technology. The low cost of CRISPR components and wide availability of requisite skills across the process suggests that smaller labs in any country are likely to be able to execute genome editing projects, and could rapidly obtain CRISPR capabilities with little effort, if they have not done so already.

## SECTION 4: GOVERNANCE FOR CRISPR

In the previous sections, we examined the tools and processes arising from the genome editing field, disaggregating the technical, social, and economic factors surrounding the adoption, utility and use cases of CRISPR in particular. This section attempts to configure these factors in a way that opens the door to policymakers’ efforts to develop ideal systems of governance for CRISPR and future genome editing technologies, and illustrate the challenges and considerations that should be weighed in these efforts.



## Why Govern CRISPR?

With a wide range of applications, and relative ease of access and use, CRISPR genome editing technology is generally considered a disruptive innovation with far-reaching potential for beneficial use in biological engineering and great scope to be misused. Without losing sight of the enormous value of CRISPR in the context of the clear benefits to the life sciences and medicine, the deliberation of oversight regimes should be conducted in good faith with traditional biotechnology user groups and all others interested in the technology, so as to develop governance structures that promote responsible deployment of genome editing, and prevent irresponsible and harmful uses of the technology, across the entire spectrum of users. Succeeding in this endeavor requires one to navigate the complexity of the field without disrupting the fertile ecosystem of technology generation and knowledge production that the tool has engendered. Notably this governance must extend to activities inside and outside of the lab, beyond expected scientific or clinical applications, in order to address the ways in which genome editing may affect society at large.

## Governance Strata

In evaluating governance options, at least three strata of genome editing activity become relevant: local, super-local or national, and international, as shown in Figure O. These are proposed as illustrative abstractions rather than absolute categories<sup>4</sup>, upon which governance structures may be superimposed.

The local stratum is concerned with the dynamics within and between scientists, genome editing users, research laboratories, institutes, company labs and commercial providers in a closed loop. Governance activity focused on this stratum might seek to understand or alter how users participate in teams that execute the genome editing process, how they access relevant materials, equipment and knowhow, and how their activities are made to meet higher level organizational requirements defined by a facility (a university or company) to meet super-local or national regulations. This stratum emphasizes localised, bottom-up approaches to organising genome editing activity, and the efforts may not extend “beyond the building.” By and large, the bulk of salient genome editing activity will be conducted within a local context, with the most important transactions of knowledge, material, and technology occurring within a particular locality. Knowledge developed in this arena may circulate between labs within a research community, via conferences or focused journals. Within this stratum, some genome editing operators or organizations may develop their own codes of conduct, working practices or internal policies, but this will not always be the case, and may also not be enforced within the context of the lab or team. Local governance does not occur in the context of diplomatic chambers or government regulator offices, but proceeds in one or more labs, and is concerned with the science proceeding smoothly.

Super-local, or national, governance is concerned with the top-down oversight of the dynamics within different local groups, including labs and companies, with one another, states, governments, or transnational organizations. It is not developed in a lab, but in the context of these government regulator offices, with super-local authority imposing restrictions, licensing,

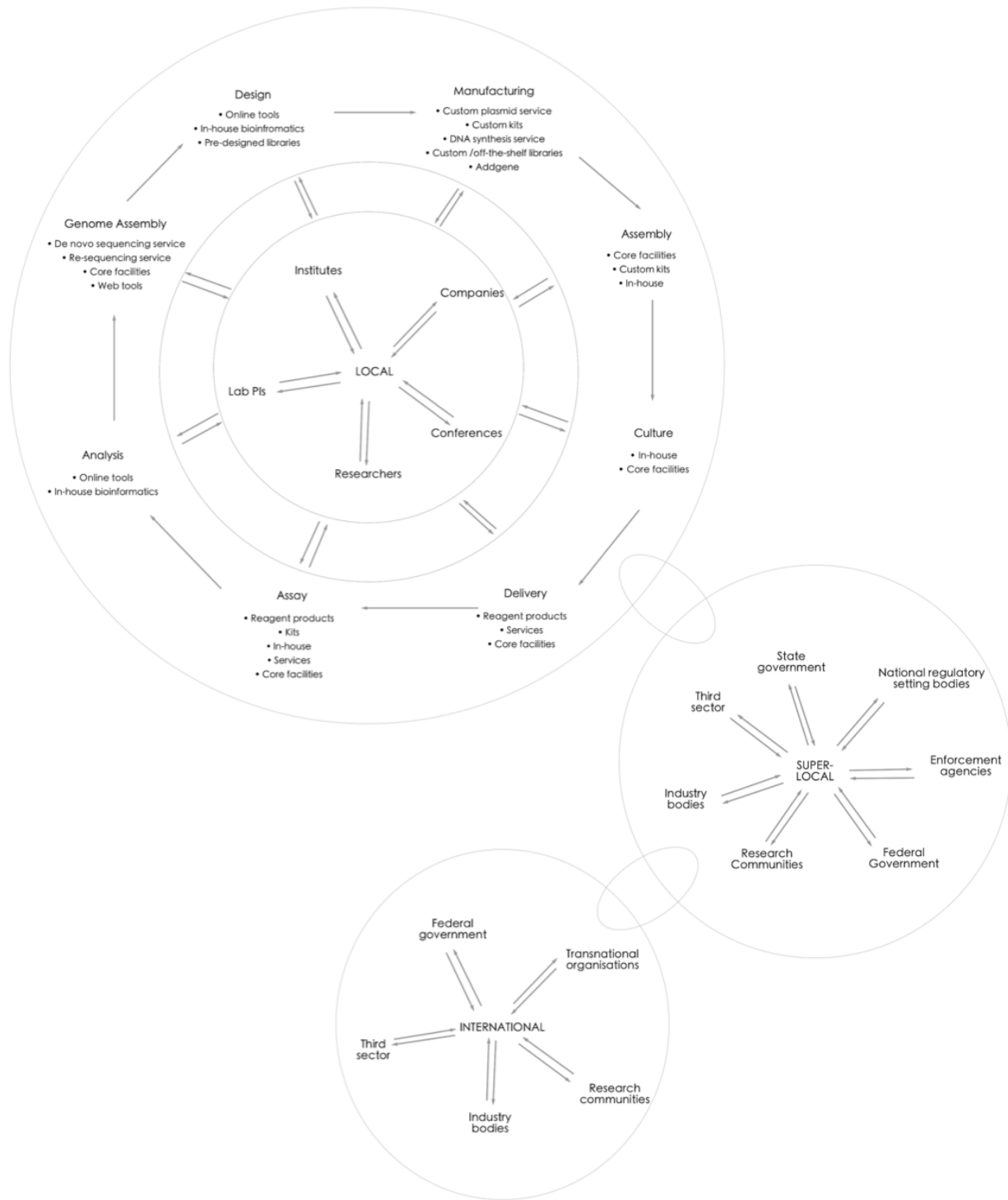
---

<sup>4</sup> An alternative approach might assign five strata revolving around knowledge sets, products, individuals, institutions or states.

controls and other mechanisms. Oversight guidelines may be imposed on citizens, local and foreign providers, researchers and research organizations operating within their borders (or selling to operators within their borders). By and large, it will be states and national agencies that seek to develop governance in this arena, likely in the form of national hard laws and regulation, which will diverge considerably in their nature, reach, and level of enforcement between different nations or organizations. It is likely that the bulk of actual governance regimes will be defined by actors within this stratum, before trickling back down to the local stratum. Notable examples include the UK House of Commons Science and Technology Select Committee on Genomics and Genome Editing, the UK Human Fertilisation and Embryology Authority's decision to approve CRISPR experiments on human embryos, and the U.S. National Institute of Standards and Technology Genome Editing Consortium, established to meet the pre-competitive needs of the emerging genome editing community. In these cases, local groups' genome editing activities inform the development of governance guidelines and decisions at this level, but decisions are handed back down to the local level from higher up.

International governance is concerned with the dynamics between different states, and emphasizes the agreements made between nations to accommodate one another's concerns regarding the kinds of research and development activity that are authorized within their sovereign territory, and how they build trust and verify one another's activities in this regard. Notable international governance efforts for genome editing include the joint summit by the Chinese Academy of Sciences, U.S. National Academy of Science, Engineering, and Medicine, and UK Royal Society, as well as discussions regarding genome editing by the Australia Group, and the United Nations Biological Weapons Convention (BWC) Meeting of Experts. Agreements made at the international level, in the form of memorandums, conventions or treaties, will ultimately have far reaching consequences at the local level, as nations will establish and implement laws at the national level, and with a mutual expectation of compliance by local operators. Despite the lofty goals of high level international governance, it largely acts to magnify and harmonize the role of super-local governance structures. In reality, it is the interactions between local user communities and super-local regulators (which might be unseen) that drive the bulk of governance parley regarding genome editing.

The key is to recognize that as the technology matures, the governance systems become more complex as they are co-developed by more groups acting at each of these different levels—oftentimes in more than one layer simultaneously. In the five years since CRISPR emerged, there have been steps at all levels in a globally distributed manner. Effective governance requires working across these levels, especially for a technology that has diffused internationally and into a variety of fields and industries.



*Figure O - Genome editing stakeholders and process interactions at local, super-local and international strata. At the local level, the outer ring represents the ideal genome editing process, with sources of wetware, hardware and software, revolving around an inner ring of user groups who may also transfer material to one another. Local processes are linked to super-local actors, which are themselves linked to international governance actors.*

## Considerations for Ideal Governance

At a minimum, the objectives of those interested in developing genome editing governance systems might: support the effective monitoring of trends in the uses and distribution of CRISPR as an emerging technology; create meaningful technology assessments that appreciate the social context of its use; understand how restrictions might interfere with technology use cases and work pipelines; forge connections between users and governance groups to promote mutual trust and understanding; restrict and record access and use of technology where appropriate; develop systems to prevent, detect, attribute, and monitor deliberate misuse or accidents; and respond to environmental release of edited organisms. In these senses, good governance seeks to understand the technology and different users' values and goals without restricting legitimate innovation, whilst also building systems that can prevent or respond to any undesirable scenario.

Some effective approaches that can be taken toward these objectives are drawn here from three schools of thought on effective governance of emerging technology, including concepts advanced by: Prof. Kathleen Vogel's Biosocial Frame (Vogel 2012), Prof. Joyce Tait and colleagues Proportional and Adaptive Governance of Innovative Technologies (PAGIT) framework (Tait et al., 2017) and Dr. Jonathan Tucker's Dual Use Decision Framework (Tucker 2012).

Vogel's biosocial frame emphasizes a detailed understanding of the skills, disciplines, organization and management, and other localized social practices, that constitute bioscience research and development. The biosocial frame examines the role of know-how and uncoded knowledge as an enabling factor for achieving a technology capability, also emphasizing how the character of technology use varies between users at different localities, and evolves along differential trajectories between these distinct groups. By incorporating the biosocial frame, one can overcome the high risk of misunderstanding a technology and its users, and reduce the chances that a technology assessment or regulatory regime is overly critical or alarmist, or does not go far enough.

According to Vogel, the alternative to the biosocial frame, the technological revolution frame, ignores the biosocial complexity to assume the worst – that anyone can use CRISPR to edit their own genes, and that technological development will likely continue in this direction until such extremes are reached. A technological revolution frame assumes that the availability of open access genome editing protocols and blackboxed kits enable actors to achieve genome engineering capability without much thought.

The biosocial frame instead articulates that understanding the explicit information of a scientific paper is not enough to achieve the same result in a lab, as this requires extensive tacit knowledge which is not uniformly distributed across the population of potential CRISPR users. A biosocial perspective supports the notion that a novice would encounter great difficulty in executing a complete and successful run of the CRISPR process, even if they had access to the most well-equipped lab in the world. Moreover, this approach also helps us to see that novices who want to execute a run of the complete CRISPR process in their own home would find it impossible without significant efforts being made to adapt the technology and process to suit this setting and this kind of user. The biosocial frame also asks us to investigate how likely all these things would be, and what additional human issues (financial, skills-based and geographic) stand in the way. Finally, this framing forces us to explore why ZFNs and TALEN technologies have

not (to our knowledge) been misused, and whether or not the factors that stood in the way of that misuse remain in place today. Regarding CRISPR, it is clear that much of the challenge and tedium of older generation GEV design and assembly has now been removed, but many other factors surrounding the other steps of the process remain bottlenecks. Notably these gaps are differentially problematic to different groups and sectors, depending on their funding, expertise and experience with genome editing.

Tait and colleague's work on the PAGIT framework proposes user-appropriate governance in the context of regulation and policy that is adaptive and proportional to the competing requirements of innovators using the same technology in entirely different sectors. In a fundamental shift away from conventional regulatory thinking, PAGIT considers how a technology or innovation can often be captured by a particular regulatory system, which will heavily modify the direction(s) that the fundamental innovation is allowed to proceed in and define the future shape and application of the technology. If new applications of the innovative technology are subsequently developed after this capture within a particular regulatory system, the innovative new application's technology development will be debilitated as it would need to conform to an inappropriate regulatory regime. It is perhaps obvious that therapeutic CRISPR applications aimed at treating cancer will need to be regulated in different ways than CRISPR-based sensors aimed at detecting disease agents in drinking water. However, if all future genome editing sectors have to deal with a clinical CRISPR regulatory system in order to solve a non-clinical CRISPR problem, this will discourage innovation in the latter arena by creating high entry costs than small biosensor startups might be able to afford, let alone make profit from. At scale, this effect would drive technology development exclusively in the direction of cancer therapeutics rather than contamination biosensors.

To this end, Tait and colleagues propose PAGIT as an adaptable governance system that requires regulators to evaluate the nuanced use cases and history of an emerging technology without making permanent decisions, instead ratcheting up from soft governance measures to harder forms as the technology matures. To achieve this, a technology readiness level (TRL) assessment is executed within the context of a sector or use case, which informs selection of a regulatory system that is proportionate to the needs of the technology at a given time, with governance updates tied to ongoing TRL assessments. PAGIT approaches are modified dependent on an innovation being incremental or disruptive to a particular business model and value chain. For disruptive technologies, governance advances from informal pre-regulatory standards to pre-regulatory guidelines in early stages, and for later stages from regulations to post-regulatory standards. For innovations that are considered incremental, PAGIT proposes that regulation should be immediately pursued (as the TRL is assumed to be already high), and notes that this is best achieved by seeking to adapt existing regulatory frameworks to new technologies, wherever possible.

Tait and colleagues highlight that the same innovation may be incremental in one sector, but disruptive in another. For instance, a CRISPR capability in the context of pharmaceutical cell line engineering projects, or agricultural biotechnology research, is somewhat incremental with well-established older generation genome editing capabilities already serving as a model to allow a larger set of actors with less specialized skills to use CRISPR at lower cost, shorter timelines and established codes of conduct ready for adaptation. The norms in pharmaceutical or agricultural genome editing are simply made more attractive to those already well-equipped to execute, or already executing, some form of genome editing process. On the other hand, some

CRISPR capabilities for instance gene drives in the context of conservation, are highly disruptive - especially given limited use of genome editing in that sector to date, which remains controversial and is largely without precedent. These examples are fairly clear-cut, and as such PAGIT provides a convenient means of assessing next steps towards governance. However, decisions about which genome editing applications are incremental or disruptive are not always clear, and are likely subject to a great many biases.

Tucker's Decision Framework on Dual Use Technology considers the risk of misuse alongside governability of new technologies, aggregating scores from these assessments to inform the development of tailored governance measures ranging from informal systems to soft law and hard law. Putting aside the risk of misuse, one is left with the factors that are most salient to governability, including: the embodiment format (hardware, software, etc.); maturity (relating to technology development pipeline and proximity to the lab or the market); convergence (number of disciplines required to employ the technology); rate of advance (of reliability, cost improvement, accuracy etc.); and international diffusion and availability of the technology.

Deploying Tucker's governability assessment to concretely define the governability of CRISPR technology quickly proves challenging. CRISPR technology has, in general, a hybrid embodiment, with low governability emerging from its intangible know-how, but medium-high governability for the required tangible goods (software and hardware), which are easier to control. Its level of convergence is generally high, with the complete CRISPR process potentially requiring an enormous range of skill sets and tacit knowledge from experienced operators, as outlined in Section 3. Whilst the required convergence is being eroded somewhat by the availability of kits and services that blackbox certain steps for those who can afford it, the current convergence remains high, rendering a low governability score due to the complexity of effectively governing so many peripheral tools and techniques associated with CRISPR. The rate of advance is high, with efforts to improve the technology proceeding along multiple directions and in various organisms, further rendering a low governability. The international diffusion has been - and continues to be - high, again rendering a low governability score in Tucker's assessment framework.

	Embodiment	Maturity	Convergence	Rate of Advance	International Diffusion
Low Governability	Intangible Information	Not Mature	High	High	High
Medium Governability	Hybrid	Very Mature	Moderate	Moderate	Moderate
High Governability	Hardware	Moderately Mature	Low	Low	Low

Figure P - "Relationships between parameters and level of governability" adapted from Tucker's governability assessment (Tucker 2012).

From the factors outlined so far we might conclude that CRISPR has a mixed-to-low governability on average. However, the maturity of the wider technology is divergent and difficult to assess at this time - some CRISPR products, notably dedicated RGEN tools and kits, are already mature and available on the market, meaning there is high governability according to the framework. However, other products or applications remain firmly in the basic research phase, suggesting there is low governability for these newer applications of the technology. Moreover, the question “how mature is CRISPR?” may be less important than “how mature is genome editing?”, or “how mature is a particular process that involves genome editing?”. One can argue that whilst CRISPR itself is not mature, or only moderately mature, and is thus highly governable, genome editing in general is very mature as at least some form of the process has been around for considerable time since the launch of first generation tools. This perspective, aligned to Tucker’s model might suggest that genome editing in general has medium governability. On the other hand, applying the broadest brush to CRISPR or genome editing to assess its governability *in general* fails to provide a meaningful path towards allocating appropriate governance measures. Rather, the governability of each step of the process, as well as the technologies and tools used within each step, ought to be examined more precisely. Finally, the governability of the individual processes that involve CRISPR- cell engineering, organism engineering, screening, and others that were neglected in this paper - also ought to be analyzed separately via Tucker’s model.

The ideal governance framework might first seek to understand how different groups are interested in the disruptive aspects of the technology, identify possible codes of conduct, and widely report these findings for uptake and adaption by others. In the initial phases, societies could be kept confident that there are at least some activities to keep an eye on technology development and use, with parties gathering data on the benefits and risks of genome editing as it emerges, with a clear understanding that governance can be ramped up to harder measures as necessary. Innovators could benefit from this system by working closely with governance actors to ensure they appreciate the nuances of the technology, and avoid bureaucratic and sub-optimal systems being built around the maturing technology. In an ideal world, governance would *always* proceed in this explicit way. However, regulatory capture oddities already exist, and they have seriously affected the course of historic biotechnologies to date, for instance the choice to treat both cell therapies and genetically modified fish as pharmaceutical drugs in the United States, rather than as novel classes of product entirely. Even now, past decisions regarding genetic engineering technology have affected contemporary CRISPR governance, the impact of which is impossible to determine. Even the most open-minded human alive today will fail to predict how CRISPR might (or should) be applied in decades to come, so achieving this ideal may not be possible especially given the nascent state of newer governance frameworks. And with mounting political pressure to govern the applications of CRISPR that are viewed as regulatory anomalies<sup>5</sup>, especially when they can be developed so easily as is the case with gene drives, the likelihood is that suboptimal security-driven decisions could well get made in the short-term at the expense of a healthy innovation ecosystem in the long-term.

---

<sup>5</sup> For more on this, refer to Weiss Evans and Palmer paper on Anomaly-handling and the politics of gene drives (Evans & Palmer 2017).

## Summarising the CRISPR Problem Space

Before exploring and overlaying governance options on top of the genome editing landscape, it is useful to assemble the issues that make CRISPR and genome editing worthy of attention, and to define the techno-social problem space more clearly than previous sections. Tucker's framework helps bring the concerning attributes into focus: the ease of use, rapid rate of development, generally low barriers to misuse, global distribution of technology and skills, democratization and deskilling of convergent tools, wide range of applications, and early stage of technology and culture development across multiple applications.

CRISPR is more tractable and simple to use than previous generations of genome editing technology. Necessary technical skills can be obtained in as little as 15 weeks, and requires no specialised or extensive capital investment beyond what may be found in a standard molecular biology lab. A user can execute basic genome editing processes using CRISPR GEVs, which come in various forms and are ostensibly readily obtained from commercial and academic sources in a partly pre-made form. Users only need to execute simple cloning procedures in order to assemble a basic experiment-ready plasmid GEV capable of knocking out a gene, and can readily purchase short oligos and reagents to do this. Kits and products are being developed by numerous providers to assist researchers. Guides and donor DNA may easily be designed using freely available or low cost software tools. Other CRISPR payload elements have been widely published and characterized, and have been shown to work in numerous model organisms, life stages, and cell cultures. Many of the drivers of simplification associated with synthetic biology (abstraction, standardization, modularity, automation, and rational design tools) are embedded in CRISPR technology. Whilst challenges remain, notably off-target issues, mosaicism, germline editing, and detection and validation of editing, these are not barriers to entry, only hurdles to elite technical achievement. Having said all this, CRISPR is not absolutely easy to use, and each application and process will present problems for anyone who lacks the explicit and tacit knowledge required to know what they are doing, and to troubleshoot issues - this point cannot be emphasized enough.

The breadth of application is high, and has boosted well-established biotechnology activities like high throughput screening and cell line engineering, and lifted new activities into the limelight, like gene drives for biocontrol. The technology can be used to reverse-engineer genotype-phenotype relationships and discover the role that genes play, but also exploit that knowledge as a forward engineering tool to generate cells and products. The wetware components can be deployed in a huge variety of GEV formats, including plasmid, virus, agrobacterium, RNA, RNP, and nanoparticle formats, providing delivery options for a variety of cells, tissues and organisms. A high rate of user innovation drives the development of new RGEN variants that allows fine-tuned control of genome expression, around which many varieties of more complex *genome engineering* are emerging. The field is indeed undergoing a renaissance, catapulting treatment for infectious and genetic disease ahead in humans, and fostering great innovation in plant and animal biotechnology.

The rate of development of CRISPR techniques is high, with significant improvements on CRISPR/Cas9 being published on at least a monthly basis, and new variants of CRISPR RGEN offering novel alternatives. Synergistic advances and ecosystem structure drive a virtuous cycle that attracts further user-innovators to the space, expanding the breadth of application as they forge new paths and processes in different models and applications, largely under an open



source banner. The availability of this low-cost, useful tool has stimulated the entrance of new players from more esoteric branches of biology. Increasing attention is paid to automating and blackboxing some (but not all) steps of the process, and commercial providers compete on price to offer relevant kits, reagent products, or services. Other aspects being improved include, but are not limited to, the range of target binding sequences, on-target efficiency, off-target activity prediction, and the number of organisms that can be modified.

CRISPR know-how is globally distributed, with the necessary wetware now found in tubes and plates in freezers in many countries, cities, universities, companies, and even homes. Protocols and test data are available from online academic literature, free commercially-produced content, and at public academic conferences where leaders report new techniques. Paywalls for codified content are outdated or ineffective and this knowledge can be gathered online at little to no cost. Older generations of genome editing technology have served as foundations for the development and use of CRISPR. The necessary genome information required to design editing strategies for different organisms is available in online databases, and the equipment required is routinely found in many biology labs, or can be acquired at low cost from secondary market providers. Acquiring the necessary tacit knowledge to design, execute, and analyze a CRISPR genome editing process, is perhaps the most difficult aspect relating to know-how, requiring on-the-job training or access to academic courses. Whilst some amateur courses and educational kits are available, they do not equip learners with the necessary skills to handle all of the technical art of the ideal genome editing process (or any specific process) defined above, which requires many years for contemporary experts to master.

Other social and legal barriers to obtaining and using CRISPR reagents are ostensibly low, and record-keeping about their use is limited. In truth, commercial and academic providers do implement some safeguards, with varying levels of customer screening or vetting enacted (see below), but there is rarely any facility- or government-enforced restriction on acquiring, maintaining, or documenting the use of CRISPR or enabling materials across the process. Activities and acquisition can easily be concealed from authorities and laboratory colleagues. There are some publicly available records, including public grant applications and mandatory registrations with national genetic modification bodies or agencies, but these vary significantly in scope and structure by territory. Restrictions on intellectual property for some clinical or commercial applications of CRISPR do not act as a meaningful barrier to entry, and only open the door for patent holders to take legal action if they are inclined to. Neither current public records nor IP systems should not be considered as barrier against use - and especially not misuse

The biological barrier to using CRISPR depends on the activity in question, and if the user intends to bridge a fundamental gap in underlying biology, or is exploiting existing knowledge. As a rule of thumb, it will be easier to use CRISPR to engineer a cell or organism with a known trait, where others have completed the work to identify and characterize that trait, than it would be to both discover and engineer the trait. To this extent, biological barriers to use CRISPR must be considered in the context of the project objective, with the number of steps involved reducing the overall chance of success. Similarly, the technical barriers will also vary considerably depending on what is required for the process. Germline human editing, for instance, presents a high equipment and skills barrier, with more substeps and complexity per step than fruit fly gene drive engineering. All things being equal, the technical barriers to genome editing are low enough to raise flags, and as for oversight and governance, it remains too

early to tell what will happen organically for the time being, highlighting a strong need to continue exploring which parts of the problem space should be prioritized.

## **Situation Report & Governance Options Across the Ideal CRISPR Process**

Rather than conduct an in-depth governability assessment of each process, as outlined in section 3, which is outside of the scope of this volume, here a broad set of practical options relating to governance of the ideal CRISPR process is combined with a description of the challenges in their implementation, in the hope that this might help guide more detailed discourse on the matter in the near future.

Looking primarily at practical governance options at the local stratum, as shown in Figure Q, one sees a number of governance mechanisms emerge across the ideal genome editing process, namely restricting or interdicting access to particular kinds of equipment, software tools (or datasets), and wetware reagents relevant to each step. These mechanisms are intended to limit the solutions that are available to users in the marketplace, and are largely artifact-centric in that governance is oriented toward top-down control of tangible goods or interactive services. Other approaches, such as licensing users, or building databases of experiments and materials holders, are concerned with controlling access on a user or case basis, for instance by deploying analytics systems into software tools that are tied to a licensing authority, or by vetting customer orders and material transfer agreements. By and large such measures, whilst seemingly obvious, are not yet in place for genome editing technology in any meaningful way, and there are serious gaps in governance for each step of the process as well as the whole, reviewed below.

Genome sequence data is freely available from public databases, with the US National Center for Biotechnology Information (NCBI) alone listing 6,479 eukaryotic (plants, animals, fungi etc.), 155,735 prokaryotic (bacteria) and 18,133 viral genomes online as of August 2018. These genomes include those of Select Agent pathogens that pose a severe threat to human, animal, or plant health. Whilst the entire set of genomes are of varied quality, many will have sufficient data to support the design of guide RNA and donor DNA for loss-of-function and gain-of-function CRISPR work. At present, there is no governance structure in place to oversee who has access to this data, which can be downloaded and imported into a dedicated genome editing design package (if that genome is not already available on a public CRISPR tool).

From a security perspective, a governance structure around genomes might ideally impose limits on the kinds of genome data that could be accessed, making certain public datasets out of bounds for a general population of users. However, this would require substantive delisting (and negotiation) efforts, where the out of bounds sequences or genome datasets are specifically blacklisted and removed from public databases. Not only would this action negatively impact legitimate research, the effort would only be as complete as our understanding of which organisms and genes are harmful. This knowledge gap is considerable, especially given the surprising ways in which apparently harmless aspects of biology can become dangerous if manipulated in a non-obvious way. Moreover, attempts to remove a genome from public access may not be effective, especially if the data remains available offline, is uploaded to the Dark Web, or if the genome of a closely-related organism remains publicly available. Indeed, it is not outside the realms of possibility that harmful sequences could be reconstituted using human or machine-learning approaches. Finally, the removal or restriction of public genome databases or



Figure Q - a schematic highlighting potential governance interventions concerning relevant materials, hardware, software and wetware reagents for each step in the CRISPR process, as executed at the local level. Red text illustrates possible governance mechanisms that can impact sources or collective use of material, equipment, or software, or interactions between groups. Not shown in this schematic is know-how, which is required across, and particular to, each step and cannot easily be controlled in this context. This schematic is intended to provide a scaffold upon which additional structures, such as positive incentive mechanisms, can be further added.

datasets would not significantly impact a user's capability to execute a *de novo* sequencing project of their own, using a sample acquired in the wild, or to hire a service provider to do this on their behalf. To this end, sequencing hardware and bioinformatics pipelines would also need to be subject to governance at the international level, a step that seems impractical or impossible.

The genome assembly and design steps are intrinsically linked in a manner that creates a notable hurdle for using CRISPR in non-human organisms. The hurdle arises from the fact that current standards for evaluating CRISPR guide design are generally based around scoring systems that are trained to select guides most relevant to human and mouse genomes. The relative paucity in publicly-funded research in other organisms means there are fewer tools dedicated to guide design in non-human organisms. Whilst an organism-specific scoring algorithm is not a necessity, its absence may complicate the rational design of guides in other genomes, and has potential to negatively impact outcomes for users working in these organisms. Likewise, the development and publication of a scoring function tuned to guide design in fungi or bacteria could positively impact outcomes for users in Select Agents in these taxa.

Genome assembly and design steps are also linked when it comes to the quality and format of the underlying dataset. Non-human and non-model genome assemblies and annotation datasets vary considerably in terms of their quality and format. Some genome datasets have been ported to highly accessible and visual genome browsers whereas others exist as fragmented low quality text-based datasets. For the latter, the variability makes it difficult to apply standardised bioinformatics pipelines to these non-standard datasets, and port them in a dedicated genome browser, unless the user has relevant bioinformatics skills. Those working on esoteric organisms would potentially have to overcome this challenge, which may require re-sequencing an organism's genome in order to obtain a superior dataset that can be ported to a genome browser ahead of guide design. Having said that, a high quality build of an annotated genome, available in a genome browser, is not strictly a necessity either - one could conceivably use a control-find shortcut function in a text-based genome file containing raw data to identify possible NGG Cas9-binding sites and select guides in this rudimentary way. Governance actors may wish to keep these issues in mind as they consider the likely workarounds to effective governance of these two steps.

Being largely software-oriented, CRISPR design tools suffer from challenges similar to those posed by genome assembly. Already there are dozens of online and offline software packages that can be obtained at no cost. A number of pre-designed virtual guide libraries are available for some organisms, including the human genome, obviating the need to design guides in some cases. Many CRISPR design tools are easy to operate, accompanied by extensive documentation or webinars that support the rapid acquisition of theoretical knowledge. Restricting access to these tools lies counter to the interests of the academic labs and companies that have developed them, although one may find success in lobbying tool operators to remove, or refrain from adding, certain genome datasets that have a high risk of being misused. Commercial operators of tools may have implemented analytics systems that monitor user activity, such as genome selection, and donor sequence design, but unless the operator has also attempted to verify the user's identity and location, so that a red flag can be associated with a real human when misuse is suspected, this data may not be helpful for guiding governance actors or law enforcement agents to a suspect.

Manufacturing and CRISPR payload assembly are primarily concerned with wetware-focused service providers delivering low value, high volume DNA synthesis services to a commodity market, as well as some manufacturers providing high value batches of pre-assembled GEVs (viruses etc.) on top of basic guide RNA manufacturing. The alternative to purchasing from these providers is to purchase DNA synthesis equipment suited to the in-house production of basic nucleic acids, which is prohibitively expensive for small labs, a fact which largely confines this kind of equipment to core facilities and service providers. Although exact numbers and distribution ranges have yet to be verified for this equipment, the centralisation of this capability, and limited distribution relative to downstream users, renders it amenable to governance intervention.

Some DNA manufacturers have tended to be ahead of the game regarding governance, with several companies participating in an superlocal and industry-led consortium, the International Gene Synthesis Consortium (IGSC), to monitor both customers and the DNA sequences they order, such that “dangerous” sequences (namely those relating to toxins and Select Agents) can be flagged and reported to law enforcement. Whilst the IGSC screening protocol is likely to flag cases where a CRISPR user has requested manufacturing of a donor DNA molecule from a harmful organism (as that sequence is likely to occur in their blacklisted database), IGSC does not appear to have any CRISPR-focused (or larger genome editing focused) oversight capability. Moreover, the current approach taken by IGSC, where sequence information directly serves as a warning light, is not at all suited to genome editing governance. To put it simply, when a user submits a 20mer guide sequence to an IGSC member, there is no contextual information (organism and gene) provided about what that guide sequence will be used to cut, leaving it to the IGSC to work out if that guide could be used to edit an organism and cause harm. Moreover, that information, even if it were collected, would not be indicative of a buyer’s intent to cause harm. Simply put, the same 20mer guide sequence could occur in so many genes, in so many organisms, that it would be difficult to assess the absolute intent of the user, and a user could likely mask their intent if they cared to. Notably this issue is most relevant to loss-of-function work, as gain-of-function work would likely require a specific harmful sequence to be manufactured, which would likely be captured by the IGSC screening protocol. Unfortunately, IGSC has yet to address the ease with which non-pathogenic coding DNA sequences (CDS) which escape current screening methods, could be synthesised and used nefariously. For instance, a CRISPR user could manufacture genes relating to ecosystem niche preference for a harmless organism, and knock these into an esoteric pest species to modify its range - an ostensibly simple genome editing procedure in which the components would escape “harmful” sequence screening, and potentially cause immense economic disruption if the procedure is realised.<sup>6</sup> This gap is especially profound in the context of target selection of gene drives, which are so simple to engineer on a molecular level, but so concerning in terms of their application, that there is a serious gap worthy of attention in the sequence screening space. Finally, IGSC membership does not encompass the entirety of DNA synthesis providers, some of whom are yet to be convinced of the business-case for implementing costly customer and sequence screening operations. As such, the best governance system that is available, has serious

---

<sup>6</sup> In reality it is unlikely that habitat range could be easily engineered in such a simple manner. The project would require deep expertise, experimental experience, and culture capabilities to work with the organism, and potentially require multiple genomic loci to be modified in parallel. The combination of these requirements would likely be a limiting factor to inexperienced researchers conducting this type of work.

gaps regarding genome editing, is not uniformly applied, and allows users to escape it by purchasing from non IGSC-members.

Crucially, a number of novel CRISPR delivery techniques, such as synthetic RNP manufacturing, lie completely outside the scope of the IGSC and DNA synthesis in general. Key providers of non-DNA GEV formats are not part of the IGSC, nor any other body that provides analogous oversight for RNPs. As such, additional loopholes appear to be emerging in the already-flimsy oversight environment for commoditized genome editing reagents, suggesting the need to either expand the IGSC mission, scope and membership, or otherwise develop a novel enterprise that can provide appropriate oversight and guidance to the emerging industry. Ideally, measures would extend to kits, libraries and other reagents, where they could be shown to be of meaningful concern.

The culture step is ostensibly well-governed, depending on the nation where the work is conducted. It is common that a national body requires a laboratory to register the culture of genetically modified organisms with a relevant authority. By and large, core facilities, laboratories and institutions, or companies will take care of this compliance routinely, and also follow the necessary codes required to support the use of particular kinds of GEV (for instance, lentivirus work will require a facility to have a BSL-2 laboratory certification, which must regularly meet the requirements of inspectors). However, these requirements are not mandated or enforced in the same way around the world, and the existence of these bodies and requirements does little to guard against clandestine GM culturing activities, especially given the ease with which culturing equipment can be acquired from secondary markets, and many reagents can be purchased without any customer or order screening. Clandestine activity may not only happen outside an authorised lab, but also within a well-equipped one unless there is strict local enforcement of an experimental schedule, or laboratory members have a strong team-oriented culture of responsibility and mutual oversight.

Culture hardware is globally diffused, meaning any attempt to restrict access would require substantive effort to locate and interdict existing capabilities. Culture wetware reagents are also so ubiquitous to general biology, that any attempt to govern access to these would also be ineffective. Any attempt would easily be worked around with the wide availability of tupperware food containers, horticulture, or animal husbandry materials to serve as makeshift alternatives to laboratory-grade options. Other than monitoring, licensing, or restricting work with particular organisms, governance of the culture step seems to be challenging and not worthy of significant attention at this time.

Delivery suffers from similar issues, with dedicated hardware, such as electroporators and microinjectors, being globally distributed and available in secondary markets, and relevant molecular biology reagents again likely to be too common to effectively control. Recalling that GEV assembly and delivery steps are intrinsically linked (as assembled GEVs are intended to be delivered in specific ways), governance of delivery might be best handled by addressing the nucleic acid manufacturing and GEV assembly steps via codes of conduct for sales.

Assays and analysis are conceptually more challenging to deal with, the question being whether or not it is even necessary to govern these steps at all. On the one hand, basic reporter assays and sequencing services would be useful to any actor attempting to verify the successful outcome of their projects, and in this sense would be helpful to control from a security perspective, by adding one final hurdle. On the other hand, the embodiment of these steps in the

form of common wetware reagents, commoditized DNA (primer synthesis), and commoditized sequencing services (much the same as DNA synthesis), means it would be difficult to impose CRISPR-centric governance or sequence screening systems upon already-ubiquitous materials, as their pertinence to a CRISPR assay, rather than some other assay, would generally not be evident to the group supplying the materials.<sup>7</sup> Governance of reporter assays would interfere with other biotechnology work, and would be disruptive to legitimate work. Analysis software could conceivably be put behind licensing walls, or monitored, but this would again require buy-in from academic labs and commercial operators, and would be against the spirit of intellectual freedom that is essential for scientific discovery.

In the case of high throughput screening (HTS) and trait discovery projects, analysis does in fact play the key role for identifying phenotype-genotype relationships from convoluted datasets, and is especially important for understanding complex traits governed by multiple genomic loci. Whilst there are software tools available, they require exceptional knowledge of bioinformatics, command line interfaces, functional genomics and troubleshooting capabilities to develop or use operational pipelines capable of handling high volumes of data. The role and complexity of robotics operations in HTS should also not be forgotten. Know-how is crucial in HTS applications, and its most likely to be found in teams operating at the elite end of the genome editing spectrum, where in-house HTS platform development is often observed for novel research enterprises, and is also likely to be the subject of intense local governance and managerial oversight.

Whilst the equipment and knowledge barriers to HTS are likely to erode in coming decades, it is perhaps unlikely that high-throughput CRISPR screening will become decentralized in the same way that CRISPR cell line engineering has. Whilst it is true that CRISPR has enabled cheaper use of screening methods, creating a space for newcomers, there can be heavy capital expenditure and recruitment burdens to set up this capability *de novo*, diminishing its convenience and pull. Further, established screening operators are likely to see the increased attention around CRISPR as a market opportunity to expand already-existing HTS operations and meet this demand. To this extent, elite screening facilities already invested in this capability may benefit from the economies of scale offered by CRISPR much more than newcomers, to such a degree that HTS newcomer entry is outcompeted by the expansion of existing facilities and service-providers. If this is the case, we might expect governance of HTS to be more easily achieved amongst a relatively small group of operators compared with the total population of those interested in HTS. On the other hand, the emergence of low-cost laboratory automation platforms like OpenTrons point to ongoing reductions in newcomer expenditure on automation that are allowing more to participate in HTS-like capabilities at lower cost than ever before. One should keep in mind that such innovations occurring in tools and processes at the periphery of genome editing can thus influence the dynamics of the market and technology use.

More generally, one should keep in mind that wherever governance can be dialled to meet a smaller group of commercial providers, rather than a larger group of end-users, the job of

---

<sup>7</sup> Some companies offer vertically integrated product lines to meet a significant portion of the genome editing process, with guide design tools, synthesis services, GEV assembly, and assay kits. It is conceivable that these companies would have an easier time of screening customers and their purchases, and recognising elements of concern when they are part of a larger order. However, vertical integration amongst biotech suppliers is not ubiquitous, with many smaller companies specialising in products suited to a single step in the genome editing process, allowing users to easily bypass any benefits offered by an integrated supplier architecture.

governance actors will be made easier. Indeed, commercial providers can not only act as sentinels, with expertise to understand the context of a customer's requirements, but also as innovators that develop novel technical approaches to implementing governance measures at the point of sale and exchange of goods. To this end, some evolution and expansion of the IGSC-style model to encompass genome editing is worthy of attention and discussion in the coming years.

## **The Changing Context of Supply**

A bird's eye view of the governance landscape for genome editing hides much of what is happening on the fringes of the genome editing community, and the difficulty of governing all supplier activities occurring outside the scope of traditional industry and research (The ODIN, 2018).

The ODIN has caught the attention of policymakers for its sale of at-home CRISPR kits, molecular biology reagents, basic equipment, human-ready plasmids, and live frogs upon which genome editing kits may be used. A key objective of The ODIN is to promote public understanding of biotechnology, and to provide facile access to the enabling tools and technologies required to execute experiments at home. Whilst The ODIN publicly states that no one should attempt to edit their own genomes, the founder controversially injected himself with CRISPR GEVs in front of a live audience, and previously launched a crowdfunding campaign to raise awareness for his company's kits. Further, by catering to (and emerging from) a biohacker philosophy, in which oversight is controversial, the company ostensibly rejects traditional governance regimes as ineffective and unnecessary, and permits its customers to escape them

On the one hand, a wider appreciation of the technology is wholly agreeable, and should be applauded, and there is surely no better way to understand a technology than to have used it and troubleshoot the procedure oneself. By widening the community of CRISPR-literate stakeholders, The ODIN arguably succeeds in ensuring there are "more eyes" watching out for misuse of the technology. On the other hand, The ODIN emphasises use of CRISPR technology in non-traditional settings that are outside the scope of almost all existing governance frameworks. The growing interest in CRISPR among citizen scientist users, who lie outside existing oversight systems, amplifies concerns of misuse by a growing, skilled-up community and raises important questions for regulators. If both a commercial provider, and its customer-base, are resistant to any form of oversight (a recognised assumption that is open to debate), how ought governance actors best approach the community that is developing around ODIN and its products? An outright ban of CRISPR outside of a laboratory risks pushing users underground, whereas permitting these activities and exchanges stokes fears and risk of misuse. The ODIN is a challenging example of what regulators and governance actors may be likely to deal with more in the future – cases with no easy answers.

Notable interactions between The ODIN and governance actors took place in March 2017, when the Bavarian Health and Food Safety Authority (LGL), which had interdicted shipments of ODIN Kits in German customs, reported that they had identified as many as five BSL-2 pathogens in the kits (LGL 2017). German authorities banned the ODIN's kits, and called on a Europe-wide ban on the product (BVL 2017). In May 2017 the European Centre for Disease Prevention and Control (ECDC) issued a statement recognising a low present risk posed by these kits, without ruling out a future ban given further distribution of those contaminated kits (ECDC



2017). Whilst this event is not a primary security risk (the pathogens are found on human skin and are generally assumed to be safe for healthy individuals), the public association of pathogens, CRISPR, and biohacking is not helpful for safeguarding biotechnology endeavours in general, serving to increase perceptions of abnormal dreaded risks, and potentially stirring regulators to view the technology as an anomaly that seems to require overbearing short-term action to solve problems today, at the expense of long-term innovation tomorrow.

This trajectory is admittedly speculative and not fixed. In an effort to belay public and regulatory concerns, some segments of the biohacker community condemned The ODIN for its disregard of normal safety procedures, and mounted their own investigation of the kits, which were tested in collaboration with the European Molecular Biology Laboratory (EMBL) using metagenomic sequencing (Trojok 2017). Preliminary results appear to support the LGL findings of BSL-2 contamination.

The self-policing activity occurring in the biohacker community highlights the presence of allies with whom governance actors might work in order to obtain advance warning about concerning activities, or otherwise pressure bad actors to alter their behaviour. Indeed, no contamination of ODIN kits have since been reported. At the same time, it can be argued that the presence of BSL-2 pathogens on a kit is by no means sufficient reason to interdict a shipment of otherwise normal molecular biology reagents - BSL-2 pathogens are commonly found on human-associated materials, and contamination in laboratory settings is entirely routine. In this sense, the incident could be regarded as a political decision by German authorities to shut down a disagreeable enterprise that seriously complicated their current governance capabilities. Indeed, German law prohibits the creation of genetically modified organisms and the German consumer protection agency BVL has warned that users of DIY biohacking kits may be fined or even imprisoned if they release a GMO into the environment (BVL 2017). Certainly, the easiest way to stop The ODIN and others like it would be to ban its products, but this is at best the right thing for the wrong reasons, and at worst a breach of personal liberty. Moreover, the search for the right reasons proves difficult, and any decision made for a non-traditional supplier like The ODIN would likely set up a complex precedent for how governance actors ought to deal with traditional suppliers selling the same or similar product. In other words, one would need to identify exactly why The ODIN selling CRISPR plasmids to citizens is more disagreeable than Addgene selling CRISPR plasmids to academics, and who has access to scientific experimentation. It should also be noted that The ODIN appears to conduct its operations in full view of the public, and any effort to dismantle it may simply push these disagreeable activities further under the radar, to a different supplier or secondary market, occluding them from the attention of regulators and governance actors entirely. As such, it may be time to harmonise and expand the concept of “legitimate suppliers,” or to otherwise identify the essential qualities that make a supplier legitimate or not, which may be more to do with the customers than the supplier itself. Identifying ways to build bridges with new communities of technology users is the ideal next step, and there are already promising steps in this direction thanks to the efforts of many in the biohacker and citizen science community, alongside those in governance.

## SECTION 5: WHAT'S NEXT?

With the impact of CRISPR likened to that of the Polymerase Chain Reaction (PCR) technique, an advancement that forever changed the trajectory of biotechnology, the field, sector and practices of biotechnology are likely to continuously be reconfigured around CRISPR technology as it finds its place in more processes and sectors.

This paper has identified numerous topics for further research, and given the expanding penetration and impact of CRISPR amongst biologists, they are briefly listed here to provide some direction for those seeking to further understand the rise of CRISPR and to develop sensible governance systems for contemporary and future genome editing technologies:

- Analysing the virtuous cycle and generative potential of CRISPR as a platform technology.
- Understanding how different user groups deviate from the ideal CRISPR process depending on the specific process they are executing, and the organism they are working with.
- Probing how widely distributed CRISPR products and services are around the world, in particular from China, which was largely absent from earlier analysis relating to the 142 press releases largely collected from US or European outlets.
- Understanding the differential technology trajectories of CRISPR based on territory and regulatory system, paying close attention to the concept of regulatory capture and its impact on genome editing use cases.
- Revisiting the subject of the maturity of CRISPR, genome editing in general and the particular processes that involve genome editing, to understand the impact of maturity on governability.
- Expanding the taxonomy of CRISPR processes defined in Section 3, figure I, to include more nuanced activities.
- Expanding and formalising concepts of the local, superlocal, and international governance strata, shown in figure O, and applying these to develop a path forward for governance at these levels.
- Developing additional schematics highlighting decision points and governance options for various genome editing processes as shown in figure Q.
- Exploring the utility of a general genome editing licensing initiative, as opposed to restrictive measures discussed in this paper, with a consideration of incentive structures to support licensing and compliance.

- Developing a deeper analysis of the suitability of the IGSC-model for governing the sales of, and access to, genome editing related reagents, including an analysis of the vendors, products, and services that may require sensible governance, and the possible models that can be followed in other sectors.
- Mapping the distribution of DNA synthesis equipment and other enabling technology.
- Identifying the qualities that make a CRISPR provider “legitimate”, and the values judgments involved in these decisions today.
- Developing a more robust exploration of governance of CRISPR through the PAGIT framework, and the realities and challenges of implementing it for different processes at a time when the technology has already emerged.

## REFERENCES

Adame V. et al. “An undergraduate laboratory class using CRISPR/Cas9 technology to mutate drosophila genes.” *Biochemistry and molecular biology education: a bimonthly publication of the International Union of Biochemistry and Molecular Biology* 2016; 44(3): pp.263–275.

[www.ncbi.nlm.nih.gov/pubmed/27009801](http://www.ncbi.nlm.nih.gov/pubmed/27009801)

Anon. Addgene: Zinc Finger Consortium: Modular Assembly Kit. Available at: <https://www.addgene.org/zfc/modular/> [Accessed August 7, 2018a].

Anon. Licensing of IP to maximize public benefit. *Broad Institute* 2016; Available at: <https://www.broadinstitute.org/node/35316> [Accessed July 13, 2018].

Anon. r/labrats - CRISPR v. TALEN v. Zinc Fingers. *reddit*. Available at: [https://www.reddit.com/r/labrats/comments/8704t5/crispr\\_v\\_talen\\_v\\_zinc\\_fingers/](https://www.reddit.com/r/labrats/comments/8704t5/crispr_v_talen_v_zinc_fingers/) [Accessed July 11, 2018b].

Bibikova M. et al. “Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases.” *Genetics* 2002; 161(3): pp.1169–1175.  
<https://www.ncbi.nlm.nih.gov/pubmed/12136019>

Boch J. et al. “Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors.” *Science* 2009; 326(5959): pp.1509–1512. <http://dx.doi.org/10.1126/science.1178811>

Boch, J. “TALEs of genome targeting.” *Nature biotechnology* 2011; 29(2): pp.135–136.  
<http://dx.doi.org/10.1038/nbt.1767>

Brinkman, E.K. et al. “Easy quantification of template-directed CRISPR/Cas9 editing.” *Nucleic acids research* 2018; 46(10): p.e58. <https://www.ncbi.nlm.nih.gov/pubmed/29538768>

Brinkman, E.K. et al. “Easy quantitative assessment of genome editing by sequence trace decomposition.” *Nucleic acids research* 2014; 42(22): p.e168.  
<https://www.ncbi.nlm.nih.gov/pubmed/25300484>

BVL. Genetic Engineering with Biology Kits: Simple but Possibly Punishable 2017.  
[https://www.bvl.bund.de/DE/06\\_Gentechnik/04\\_Fachmeldungen/2017/2017\\_01\\_25\\_DIY-Kits.html](https://www.bvl.bund.de/DE/06_Gentechnik/04_Fachmeldungen/2017/2017_01_25_DIY-Kits.html)

Carlson, D.F. et al. “Efficient TALEN-mediated gene knockout in livestock.” *Proceedings of the National Academy of Sciences of the United States of America* 2012; 109(43): pp.17382–17387. <https://www.ncbi.nlm.nih.gov/pubmed/23027955>

Cermak, T. et al. “Efficient design and assembly of custom TALEN and other TAL effector based constructs for DNA targeting.” *Nucleic acids research* 2011; 39(12): p.e82.  
[www.ncbi.nlm.nih.gov/pubmed/21493687](http://www.ncbi.nlm.nih.gov/pubmed/21493687)

Chen, J.S. et al. “CRISPR-Cas12a target binding unleashes single-stranded DNase activity.” 2017; Available at: <http://dx.doi.org/10.1101/226993>.

Christian, M. et al. “Targeting DNA Double-Strand Breaks with TAL Effector Nucleases.” *Genetics* 2010; 186(2): pp.757–761. <http://dx.doi.org/10.1534/genetics.110.120717>

Cohen, J. “One of our reporters tried to do CRISPR. He failed miserably.” *Science* 2016; Available at: <http://dx.doi.org/10.1126/science.aal0335>

Cohen, S.N. et al. “Construction of biologically functional bacterial plasmids in vitro.” *Proceedings of the National Academy of Sciences of the United States of America* 1973; 70(11): pp.3240–3244. <https://www.ncbi.nlm.nih.gov/pubmed/4594039>

Cong, L. et al. “Multiplex genome engineering using CRISPR/Cas systems.” *Science* 2013; 339(6121): pp.819–823. <https://www.ncbi.nlm.nih.gov/pubmed/23287718>

Doench, J.G. et al. “Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.” *Nature biotechnology* 2016; 34(2): pp.184–191. <https://www.ncbi.nlm.nih.gov/pubmed/26780180>

Doyon, Y. et al. “Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases.” *Nature biotechnology* 2008; 26(6): pp.702–708. <https://www.ncbi.nlm.nih.gov/pubmed/18500334>

Durai, S. et al. “Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells.” *Nucleic acids research* 2005; 33(18): pp.5978–5990. <https://www.ncbi.nlm.nih.gov/pubmed/16251401>

East-Seletsky, A. et al. “Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.” *Nature* 2016; 538(7624): pp.270–273. <http://dx.doi.org/10.1038/nature19802>

European Centre for Disease. *Risk Related to the Use of ‘do-It-Yourself’ CRISPR-Associated Gene Engineering Kit Contaminated with Pathogenic Bacteria*. Stockholm:ECDC; 2017. [https://ecdc.europa.eu/sites/portal/files/documents/2-May-2017-RRA\\_CRISPR-kit-w-pathogenic-bacteria\\_2.pdf](https://ecdc.europa.eu/sites/portal/files/documents/2-May-2017-RRA_CRISPR-kit-w-pathogenic-bacteria_2.pdf)

Evans, S.W. & Palmer, M.J. “Anomaly handling and the politics of gene drives.” *Journal of Responsible Innovation* 2017; 5(sup1): pp.S223–S242. <http://dx.doi.org/10.1080/23299460.2017.1407911>

Frock, R.L. et al. “Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases.” *Nature biotechnology* 2015; 33(2): pp.179–186. <https://www.ncbi.nlm.nih.gov/pubmed/25503383>

Fu, Y. et al. “High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.” *Nature biotechnology* 2013; 31(9): pp.822–826.

<https://www.ncbi.nlm.nih.gov/pubmed/23792628>

Gabriel, R. et al. “An unbiased genome-wide analysis of zinc-finger nuclease specificity.” *Nature biotechnology* 2011; 29(9): pp.816–823.

<https://www.ncbi.nlm.nih.gov/pubmed/21822255>

Gaj, T., Gersbach, C.A. & Barbas, C.F., 3<sup>rd</sup>. “ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.” *Trends in biotechnology* 2013; 31(7): pp.397–405.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3694601>

Geurts, A.M. et al. “Knockout rats via embryo microinjection of zinc-finger nucleases.” *Science* 2009; 325(5939): p.433. <https://www.ncbi.nlm.nih.gov/pubmed/19628861>

Gootenberg, J.S. et al. “Nucleic acid detection with CRISPR-Cas13a/C2c2.” *Science* 2017; 356(6336): pp.438–442. <https://www.ncbi.nlm.nih.gov/pubmed/28408723>

Grav, L.M. et al. “Application of CRISPR/Cas9 Genome Editing to Improve Recombinant Protein Production in CHO Cells.” In *Methods in Molecular Biology* 2017; pp. 101-118. [http://dx.doi.org/10.1007/978-1-4939-6972-2\\_7](http://dx.doi.org/10.1007/978-1-4939-6972-2_7)

Greisman, H.A. & Pabo, C.O. “A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites.” *Science* 1997; 275(5300): pp.657–661.

<https://www.ncbi.nlm.nih.gov/pubmed/9005850>

He, Z. et al. “Comparison of CRISPR/Cas9 and TALENs on editing an integrated EGFP gene in the genome of HEK293FT cells.” *SpringerPlus* 2016; 5(1).

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4916124/>

Hough, S.H. et al. “Guide Picker is a comprehensive design tool for visualizing and selecting guides for CRISPR experiments.” *BMC bioinformatics* 2017; 18(1): p.167.

<https://www.ncbi.nlm.nih.gov/pubmed/28288556>

Iyer, V. et al. “No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice.” *PLoS genetics* 2018; 14(7): p.e1007503.

<https://www.ncbi.nlm.nih.gov/pubmed/29985941>

Jackson, A.L. et al. “Expression profiling reveals off-target gene regulation by RNAi. *Nature biotechnology* 2003; 21(6): pp.635–637. <https://www.ncbi.nlm.nih.gov/pubmed/12754523>

Jackson, A.L. et al. “Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity.” *RNA* 2006; 12(7): pp.1179–1187.  
<https://www.ncbi.nlm.nih.gov/pubmed/16682560>

Jackson, A.L. & Linsley, P.S. “Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application.” *Nature reviews. Drug discovery* 2010; 9(1): pp.57–67. <https://www.ncbi.nlm.nih.gov/pubmed/20043028>

Jinek, M. et al. “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” *Science* 2012a; 337(6096): pp.816–821.  
<https://www.ncbi.nlm.nih.gov/pubmed/22745249>

Jinek, M. et al. “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” *Science* 2012b; 337(6096): pp.816–821.  
<https://www.ncbi.nlm.nih.gov/pubmed/22745249>

Joung, J.K., Ramm, E.I. & Pabo, C.O. “A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions.” *Proceedings of the National Academy of Sciences of the United States of America* 2000; 97(13): pp.7382–7387.  
<https://www.ncbi.nlm.nih.gov/pubmed/10852947>

Joung, J.K., Voytas, D.F. & Kamens, J. “Accelerating research through reagent repositories: the genome editing example.” *Genome biology* 2015; 16: p.255.  
<https://www.ncbi.nlm.nih.gov/pubmed/26585970>

Katigbak, A. et al. “Inducible Genome Editing with Conditional CRISPR/Cas9 Mice.” *G3* 2018; 8(5): pp.1627–1635. <https://www.ncbi.nlm.nih.gov/pubmed/29519936>

Kim, D. et al. “Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells.” *Nature methods* 2015; 12(3): pp.237–43, 1 p. following 243.  
<https://www.ncbi.nlm.nih.gov/pubmed/25664545>

Kim, S.T. et al. “Response to “Unexpected mutations after CRISPR–Cas9 editing in vivo.” *Nature methods* 2018; 15(4): pp.239–240. <http://dx.doi.org/10.1038/nmeth.4554>

Kim, Y.G., Cha, J. & Chandrasegaran, S. “Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.” *Proceedings of the National Academy of Sciences* 1996; 93(3): pp.1156–1160. <http://dx.doi.org/10.1073/pnas.93.3.1156>

Kleinstiver, B.P. et al. “High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects.” *Nature* 2016; 529(7587): pp.490–495.  
<http://dx.doi.org/10.1038/nature16526>

Klug, A. “The discovery of zinc fingers and their applications in gene regulation and genome manipulation.” *Annual review of biochemistry* 2010; 79: pp.213–231.

<https://www.ncbi.nlm.nih.gov/pubmed/20192761>

Klug, A., Miller, J. & McLACHLAN, A.D. “Repetitive Zn<sup>2+</sup>-binding domains in the protein transcription factor IIIA from *Xenopus* Oocytes.” *Biochemical Society transactions* 1986; 14(2): pp.221.2–221. <http://dx.doi.org/10.1042/bst0140221a>

Koo, T., Lee, J. & Kim, J.-S. “Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9.” *Molecules and cells* 2015; 38(6): p.475.

<https://www.ncbi.nlm.nih.gov/pubmed/25985872>

Lareau, C.A. et al. “Response to “Unexpected mutations after CRISPR–Cas9 editing in vivo.” *Nature methods* 2018; 15(4): pp.238–239. <http://dx.doi.org/10.1038/nmeth.4541>

Lescarbeau, R.M. et al. “Response to ‘Unexpected mutations after CRISPR–Cas9 editing in vivo.’” *Nature methods* 2018; 15(4): pp.237–237.

<http://dx.doi.org/10.1038/nmeth.4553>

LGL. *Potential Risk due to the Use of the Do-It-Yourself Genetic engIneering Kit Fo the Company Odin* 2017. <https://www.lgl.bayern.de/presse/detailansicht.htm?tid=680089>.

Li, T. et al. “High-efficiency TALEN-based gene editing produces disease-resistant rice.” *Nature biotechnology* 2012; 30(5): pp.390–392. <https://www.ncbi.nlm.nih.gov/pubmed/22565958>

Liu, Q. et al. “Design of polydactyl zinc-finger proteins for unique addressing within complex genomes.” *Proceedings of the National Academy of Sciences of the United States of America* 1997; 94(11): pp.5525–5530. <https://www.ncbi.nlm.nih.gov/pubmed/9159105>

Li, W. et al. “MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.” *Genome biology* 2014; 15(12): p.554.

<https://www.ncbi.nlm.nih.gov/pubmed/25476604>

Lo, C.A., Greben, A.W. & Chen, B.E. “Generating stable cell lines with quantifiable protein production using CRISPR/Cas9-mediated knock-in.” *BioTechniques* 2017; 62(4).

<http://dx.doi.org/10.2144/000114534>

Lu, J. et al. “Multimode drug inducible CRISPR/Cas9 devices for transcriptional activation and genome editing.” *Nucleic acids research* 2018; 46(5): p.e25.

<https://www.ncbi.nlm.nih.gov/pubmed/29237052>

Maduro, M. “Faculty of 1000 evaluation for Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells.” *F1000 – Post-publication peer review of the biomedical literature* 2006.

<http://dx.doi.org/10.3410/f.1046846.496870>



Maeder, M.L. et al. “Oligomerized pool engineering (OPEN): an “open-source” protocol for making customized zinc-finger arrays.” *Nature protocols* 2009; 4(10): pp.1471–1501.  
<https://www.ncbi.nlm.nih.gov/pubmed/19798082>

Maeder, M.L. et al. “Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification.” *Molecular cell* 2008; 31(2): pp.294–301.  
<https://www.ncbi.nlm.nih.gov/pubmed/18657511>

McKenna, A. & Shendure, J. “FlashFry: a fast and flexible tool for large-scale CRISPR target design.” *BMC biology* 2018; 16(1): p.74. <https://www.ncbi.nlm.nih.gov/pubmed/29976198>

Mojica, F.J., Juez, G. & Rodríguez-Valera, F. “Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites.” *Molecular microbiology* 1993; 9(3): pp.613–621. <https://www.ncbi.nlm.nih.gov/pubmed/8412707>

Montague, T.G. et al. “CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing.” *Nucleic acids research* 2014; 42(W1): pp.W401–W407.  
<http://dx.doi.org/10.1093/nar/gku410>

Moore, R., Chandrabhas, A. & Bleris, L. “Transcription activator-like effectors: a toolkit for synthetic biology.” *ACS synthetic biology* 2014; 3(10): pp.708–716.  
<https://www.ncbi.nlm.nih.gov/pubmed/24933470>

Morbitzer, R. et al. “Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors.” *Proceedings of the National Academy of Sciences of the United States of America* 2010; 107(50): pp.21617–21622.  
<https://www.ncbi.nlm.nih.gov/pubmed/21106758>

Niu, D. et al. “Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9.” *Science* 2017; 357(6357): pp.1303–1307. <https://www.ncbi.nlm.nih.gov/pubmed/28798043>

Niu, N. & Wang, L. “In vitro human cell line models to predict clinical response to anticancer drugs.” *Pharmacogenomics* 2015; 16(3): pp.273–285.  
<https://www.ncbi.nlm.nih.gov/pubmed/25712190>

Novak B, Maloney T. “Advancing a New Toolkit for Conservation: From Science to Policy.” *The CRISPR Journal* 2018; Available at: <http://dx.doi.org/10.1089/crispr.2017.0019>

Nutter, L.M.J. et al. “Response to ‘Unexpected mutations after CRISPR–Cas9 editing in vivo.’” *Nature methods* 2018; 15(4): pp.235–236. <http://dx.doi.org/10.1038/nmeth.4559>

Pattanayak, V. et al. “Revealing off-target cleavage specificities of zinc-finger

- nucleases by in vitro selection.” *Nature methods* 2011; 8(9): pp.765–770.  
<http://dx.doi.org/10.1038/nmeth.1670>
- Ramirez, C.L. et al. “Unexpected failure rates for modular assembly of engineered zinc fingers.” *Nature methods* 2008; 5(5): pp.374–375.  
<https://www.ncbi.nlm.nih.gov/pubmed/18446154>
- Ran, F.A. et al. “Genome engineering using the CRISPR-Cas9 system.” *Nature protocols* 2013; 8(11): pp.2281–2308. <https://www.ncbi.nlm.nih.gov/pubmed/24157548>
- Ran, F.A. et al. “In vivo genome editing using Staphylococcus aureus Cas9.” *Nature* 2015; 520(7546): pp.186–191. <https://www.ncbi.nlm.nih.gov/pubmed/25830891>
- Rothstein, R.J. “One-Step Gene Disruption in Yeast.” In *Recombinant DNA Methodology* 1989; pp. 331–340. <http://dx.doi.org/10.1016/b978-0-12-765560-4.50024-1>
- Sander, J.D. et al. “Targeted gene disruption in somatic zebrafish cells using engineered TALENs.” *Nature biotechnology* 2011; 29(8): pp.697–698.  
<https://www.ncbi.nlm.nih.gov/pubmed/21822241>
- Schaefer, K.A. et al. “Retraction: Unexpected mutations after CRISPR–Cas9 editing in vivo.” *Nature methods* 2018; 15(5): pp.394–394. <http://dx.doi.org/10.1038/nmeth0518-394a>
- Schaefer, K.A. et al. “Unexpected mutations after CRISPR–Cas9 editing in vivo.” *Nature methods* 2017; 14(6): pp.547–548. <http://dx.doi.org/10.1038/nmeth.4293>
- Scherer, S. & Davis, R.W. “Replacement of chromosome segments with altered DNA sequences constructed in vitro.” *Proceedings of the National Academy of Sciences of the United States of America* 1979; 76(10): pp.4951–4955.  
<https://www.ncbi.nlm.nih.gov/pubmed/388424>
- Shalem, O. et al. “Genome-scale CRISPR-Cas9 knockout screening in human cells.” *Science* 2014; 343(6166): pp.84–87. <https://www.ncbi.nlm.nih.gov/pubmed/24336571>
- Sherkow, J.S. “The CRISPR Patent Landscape: Past, Present, and Future.” *The CRISPR Journal* 2018; 1(1): pp.5–9. <http://dx.doi.org/10.1089/crispr.2017.0013>
- Siwo, G.H. “The Global State of Genome Editing” 2018; Available at:  
<http://dx.doi.org/10.1101/341198>.
- Slaymaker, I.M. et al. “Rationally engineered Cas9 nucleases with improved specificity.” *Science* 2016; 351(6268): pp.84–88. <https://www.ncbi.nlm.nih.gov/pubmed/26628643>
- Smithies, O. et al. “Insertion of DNA sequences into the human chromosomal beta globin locus by homologous recombination.” *Nature* 1985; 317(6034): pp.230–234.

<https://www.ncbi.nlm.nih.gov/pubmed/2995814>

Stemmer, M. et al. “CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool.” *PloS one* 2015; 10(4): p.e0124633.

<http://dx.doi.org/10.1371/journal.pone.0124633>

Tait, J. et al. *A FRAMEWORK TO GUIDE POLICY AND REGULATORY DECISION MAKING*, Innogen Institute 2017; Available at: [https://www.innogen.ac.uk/downloads/FrameworkReport-Final\\_170717.pdf](https://www.innogen.ac.uk/downloads/FrameworkReport-Final_170717.pdf)

Tesson, L. et al. “Knockout rats generated by embryo microinjection of TALENs.” *Nature biotechnology* 2011; 29(8): pp.695–696. <https://www.ncbi.nlm.nih.gov/pubmed/21822240>

The ODIN. “About Us.” 2018. <http://www.the-odin.com/about-us/>

Thomas, K.R., Folger, K.R. & Capecchi, M.R. “High frequency targeting of genes to specific sites in the mammalian genome.” *Cell* 1986; 44(3): pp.419–428.

<https://www.ncbi.nlm.nih.gov/pubmed/3002636>

Thompson, N. & Zyontz, S. “Who Tries (and Who Succeeds) in Staying at the Forefront of Science: Evidence from the DNA-Editing Technology, CRISPR.” 2017; Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3073227](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3073227) [Accessed July 13, 2018].

Trojok, R. Letter to Edward Perello. November 2017.

Tsai, S.Q. “Challenges for Sensitive Quantification of Gene Editing “Off-Target” Activity.” *Small Methods* 2017; 1(1-2): p.1600062. <http://dx.doi.org/10.1002/smtd.201600062>

Tsai, S.Q. et al. “GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.” *Nature biotechnology* 2015; 33(2): pp.187–197.

<https://www.ncbi.nlm.nih.gov/pubmed/25513782>

Tucker, J.B. *Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies*, MIT Press; 2012.

Veres, A. et al. “Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing.” *Cell stem cell* 2014; 15(1): pp.27–30. <https://www.ncbi.nlm.nih.gov/pubmed/24996167>

Vogel, K.M. *Phantom Menace or Looming Danger?: A New Framework for Assessing Bioweapons Threats*; 2012. <http://dx.doi.org/10.7312/lako14606-009>

Wilson, C.J. et al. “Response to “Unexpected mutations after CRISPR–Cas9 editing in vivo.”” *Nature methods* 2018; 15(4): pp.236–237. <http://dx.doi.org/10.1038/nmeth.4552>

Wood, A.J. et al. “Targeted genome editing across species using ZFNs and TALENs.” *Science* 2011; 333(6040): p.307. <https://www.ncbi.nlm.nih.gov/pubmed/21700836>

Wright, D.A. et al. “Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly.” *Nature protocols* 2006; 1(3): pp.1637–1652.  
<https://www.ncbi.nlm.nih.gov/pubmed/17406455>

Wrighton, K.H. “Genetic engineering: Expanding the reach of Cas9.” *Nature reviews. Genetics* 2018; 19(5): pp.250–251. <https://www.ncbi.nlm.nih.gov/pubmed/29527014>

Yang, L. et al. “Genome-wide inactivation of porcine endogenous retroviruses (PERVs).” *Science* 2015; 350(6264): pp.1101–1104.  
<http://dx.doi.org/10.1126/science.aad1191>

Young, J.J. et al. “Efficient targeted gene disruption in the soma and germ line of the frog *Xenopus tropicalis* using engineered zinc-finger nucleases.” *Proceedings of the National Academy of Sciences of the United States of America* 2011; 108(17): pp.7052–7057.  
<https://www.ncbi.nlm.nih.gov/pubmed/21471457>

Zhang, F. et al. “Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription.” *Nature biotechnology* 2011; 29(2): pp.149–153.  
<http://dx.doi.org/10.1038/nbt.1775>

Zhang, F. et al. “High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases.” *Proceedings of the National Academy of Sciences of the United States of America* 2010; 107(26): pp.12028–12033. <https://www.ncbi.nlm.nih.gov/pubmed/20508152>

Zuo, X., Fan, C. & Chen, H.Y. “Biosensing: CRISPR-powered diagnostics.” *Nature Biomedical Engineering* 2017; 1(6): p.0091. <http://dx.doi.org/10.1038/s41551-017-0091>