BIOTECH STUDENT ASSOCIATION

SYNTHETIC BIOLOGY FEATURE

THE SYNBIO REVOLUTION

BIOTECH RESEARCH CLUB

BY THE
BIOTECH RESEARCH CLUB

IN THIS ISSUE

Synthetic Biology Goes Cell-Free

Burhan Sabuwala

Introduction

Life has evolved on earth for about 3.5 billion years and has survived all odds ranging from volcano eruption to asteroid impact events. This has been possible because life has the ability to reproduce, evolve and adapt to its surroundings. Now, with recent discoveries and the available modern technologies, we can manipulate organisms to do something that suits our purpose. For example, we have been using microorganisms to produce alcohol, bread, curd, etc.

However, we now have technologies to cut and paste DNA. We can also synthesize new DNA which can be used in making novel proteins that can perform novel functions. It can also be used to create genetic circuits, analogous to electrical circuits, which can perform various functions like oscillation, switch-like behaviour, memory storage, etc. Engineering life to create something that is not found in nature and can have a useful purpose is synthetic biology.

Lots of research has been done in Synthetic Biology and many of its applications have been shown in a variety of fields such as Detection of Heavy Metals using biosensors, manufacturing of certain macromolecules and polymers using microorganisms or cell cultures, diagnostics, bioremediation, producing cleaner and cheaper energy, and many more. However, not many of these inventions have come out of the lab. The major reason for this is the problem of biosafety. The organism that is modified to do a certain task cannot be released in the environment as we do not know how it will interact with the environment outside the lab.

Cell-free systems

Cell-free systems (CFSs) are basically all components of the cell that are required for our function of interest without a cell wall. For example, if we want overexpression of a protein, we can have DNA coding for the protein, RNA polymerase to form mRNA, ribosomes, tRNA and amino acids in a solution. Start-ups like GreenLight and Sutro have shown low-cost production of RNA and protein from cell-free systems which are quite safe for use. Another example of cell-free systems is PCR where we can amplify a specific DNA sequence without using any cells in a solution. This process is very quick.

This overcomes the issue of biosafety as the components cannot multiply themselves without cells. A simple process of filtering gives a bio-safe format for distribution outside the lab. Moreover, it allows for quick characterization of genetic circuits that we intend to build. It also minimizes cross-talk (interaction of our module of interest with other modules which leads to unexpected outcomes) as there is a minimum number of components in the system, unlike in cells. This also ensures better control over the process of the system, a privilege which we usually don't have while working with cells.

Another great advantage of cell-free systems is easy storage and transport. After a process called freeze-drying which is basically drying out all water at low temperature, we can store the contents for over a year and even transport or distribute it.

Paper-Based Diagnostics

Many groups across the globe have been working on paper-diagnostics. It is similar to using a pH paper to detect pH. The only change is you can use it for many more purposes. The paper can have genetic circuits that get activated upon exposure to water or a buffer and can be used to detect Mercury contamination or the presence of a virus (such as Zika, Dengue or Nipah). These paper-based diagnostics would be cheap and bio-safe to be used on the field and would be especially useful in case of an epidemic.

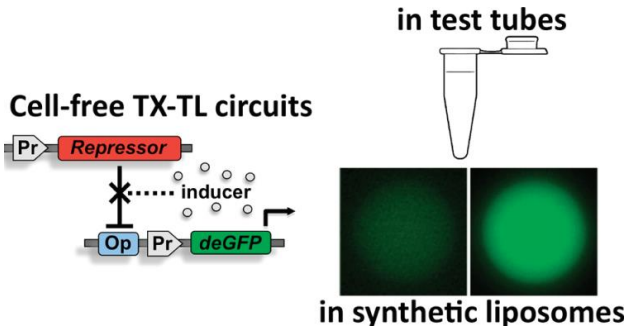
Paper-based methods are also very relevant to agriculture which is usually decentralized and access to lab testing is difficult. Paper-based synthetic biology methods can be developed to characterize soil nutrients or plant diseases. Such methods can be very beneficial to agriculture.

Production of Therapeutics

Cell-free systems have been useful for producing many peptide-based therapeutics such as vaccines, antimicrobial peptides, some antibiotics and even safe production of bacteriophages that can be useful for bacteriophage-based therapeutics. Many proteins require post-translational modifications such as phosphorylation that activates the protein, most of which are shown to be achievable in CFS.

CFS also allows for the protection of intellectual property as codon tables can be easily modified. This allows for easy encryption of intellectual property.

CFS as a Platform for Discovery

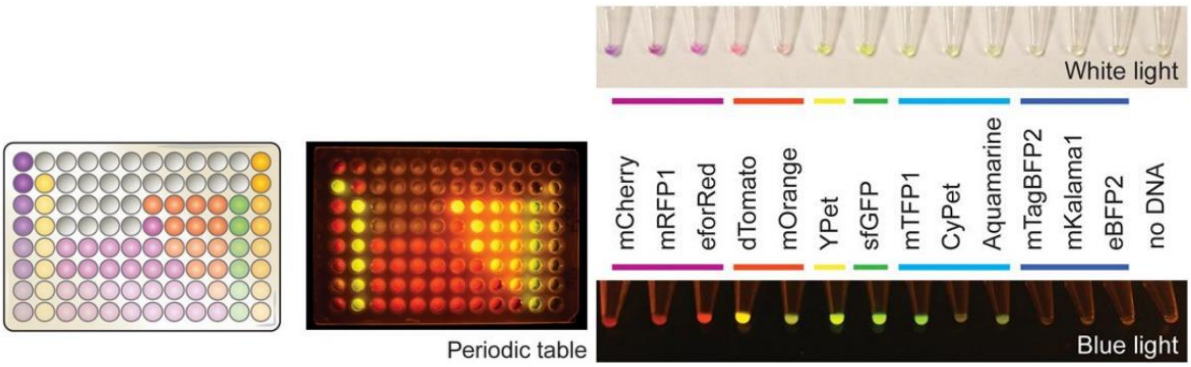


CFSs are easy to work with compared to microorganisms or cell lines when it comes to protein engineering as it is easy to crystallize. CFSs have been very useful in metabolic engineering and several groups are able to put long pathways in CFSs and also scale up to process to about 100 litres. Liposomes are enclosed lipid bilayer spheres. Genetic circuits can be enclosed in liposomes and can be prototyped individually in a single system. Membrane proteins can be easily studied on cell-free systems containing liposomes, emulsions, micelles, or oil-water interface.

CFSs can also be used for efficient energy storage. However, the research on this is still in its nascent stages. Maltodextrose based batteries have much higher energy density compared to even lithium-ion batteries.

Role of CFS in Education

CFS can be extensively used in education and spreading awareness about synthetic biology. Since it is bio-safe, it can be taken out of the lab to schools and education centres and some live experiments can be shown.



CFS technology has immense potential because of its bio-safe format, lack of a cell wall allowing for rapid prototyping, and better control over the system. It has immense potential in various areas ranging from diagnostics, agriculture, peptide-based therapeutics, metabolic engineering, bioremediation, energy storage, and spreading awareness about synthetic biology.

Reference

Tinafar A, Jaenes K, Pardee K. Synthetic Biology Goes Cell-Free. BMC Biol. 2019 Aug 8;17(1):64. doi: 10.1186/s12915-019-0685-x.

Molecular Digital Data Storage Using DNA

G. Prashant

"I am inspired by the biological phenomena in which chemical forces are used in a repetitious fashion to produce all kinds of weird effects." - Richard Feynman

Today, there are a variety of devices which we can use to store digital data, like text, photos and videos. Well-known examples include USB flash drives, hard disks, CDs, DVDs and Blu-Rays. We are also aware of a few obsolete storage devices like floppy disks and radio tapes which we do not use anymore. We are satisfied with the technology we have at the moment in terms of storing large volumes of digital data, and they have made fast-paced progress. Nevertheless, do we have to think of finding an alternative and a much efficient way to store data? Unfortunately, it turns out that these devices are approaching their density limits. In other words, the amount of data that these devices can store per unit volume are limited and fixed. As data and information are being generated in an exponential manner, it is expected that today's storage devices might also be out-dated soon. The information stored on them can only last for a few decades. This brings us to a situation where we have to hunt for a novel way to store large amounts of data for a very long period of time.

What's so special about DNA?

DNA, or deoxyribonucleic acid, as we know is the genetic material present in most living organisms. They are double-helical structures consisting of four nucleotide bases - Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) - along with sugar molecules and phosphate groups. The half-life of DNA is around 500 years in appropriate conditions, which means that it takes 500 years for just half of its bonds to break, making it highly stable. Interestingly, extensive research is taking place on using DNA to store digital data. Data is stored in electronic devices in the form of binary digits, 0 and 1. A simple idea is to map the sequence of binary digits to nucleotide sequences containing the four bases. One way to implement this is to map 2-bit sequences to the four nucleotides, as shown below.

00	**01**	**10**	**11**
A	**C**	**G**	**T**

In this way, we can convert large sequences of binary digits to sequences of nucleotide bases.

A brief history of DNA data storage

The concept of using DNA for data storage dates back to mid-1960s when the Soviet physicist Mikhail Neiman and the mathematician Norbert Wiener expressed their ideas of storing and retrieving digital data at the molecular level. In 1988, an artist named Joe Davis was able to encode a 35-bit image of an ancient Germanic rune representing life and the female Earth in a DNA sequence, which was then inserted into *E. coli*. In 1999, secret messages were hidden in DNA microdots on paper.

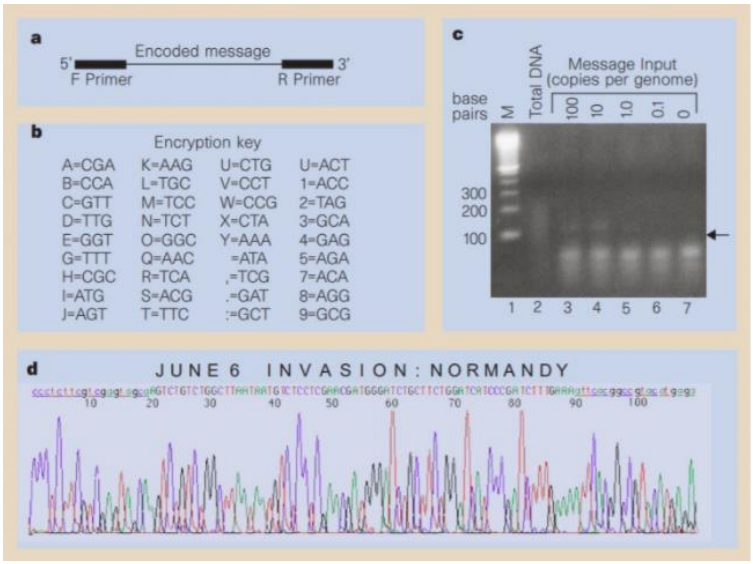


Figure 1: Genomic steganography: Secret messages encoded in DNA Source: https://www.nature.com/articles/21092

Following these remarkable attempts, several trials have been performed to improve the capacity and efficiency of DNA-based storage devices. More recently, in 2019, scientists were able to encode all of Wikipedia, totalling 16 GB, onto DNA strands.

A broad overview of the steps required

The four major steps involved in DNA data storage are: write, store, retrieve and read.

Writing and Storing

As mentioned previously, this step involves mapping strings of bits into DNA sequences by making use of computational algorithms. The binary sequences are usually broken down into smaller chunks and indices are added to each chunk to enable reassembly. The resulting DNA sequences are then synthesised. There are a variety of ways to synthesise DNA chemically. The most commonly used method is phosphoramidite-based oligonucleotide synthesis. The synthesised DNA sequences are stored in a library of DNA pools.

Retrieving and Reading

If we have to retrieve a particular data item from the library of DNA pools, we have to make sure that we choose that specific item among others rather than reading all the available data. This principle is called random access and is widely incorporated in computer memory. This can also be achieved in DNA storage by making use of PCR (Polymerase Chain Reaction) associated primers or magnetic bead extraction. Once the required data item is identified, we have to sequence it and convert it back to binary format. Again, there are a variety of ways to sequence DNA. Next-Generation Sequencing (NGS) is a massively parallel sequencing technique which is widely used on a commercial scale. Other methods include Sanger sequencing and Nanopore sequencing.

What about errors?

Both sequencing and synthesis are prone to errors like substitution, insertion and deletion of bases. In order to address these issues, error-correcting codes are implemented during the process of encoding. For example, redundant information is added to increase the probability that the original information can be retrieved even in the presence of errors and missing data. There are several ways to incorporate redundant information. Reed-Solomon codes are popular error-correcting codes that are also used to correct errors in present digital storage devices.

Storing data inside living cells!

Advancements in the field of synthetic biology have led to the construction of complicated synthetic genetic networks that can perform logical and state-dependent functions. Recording and storing of data can be done inside cells by using either a recombinase-based approach or CRISPR-based approaches. These systems can be used as biological recording devices which are more suited to collecting new data.

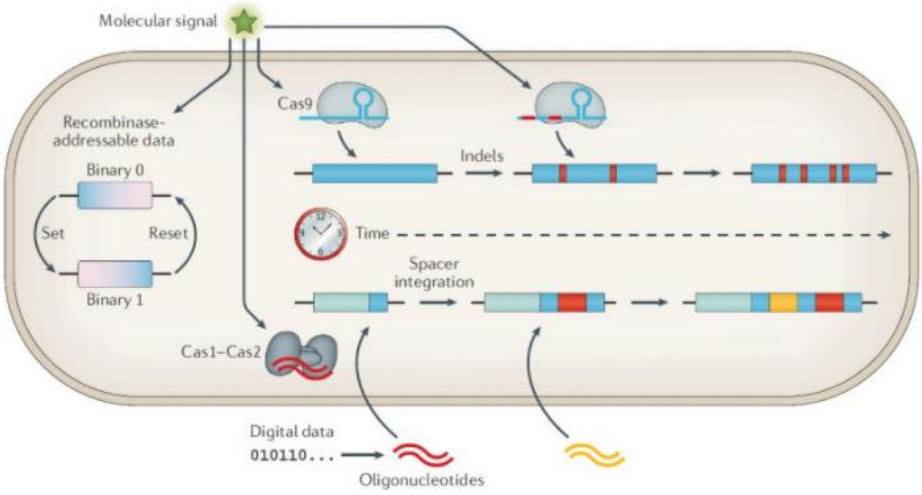


Figure 2: Storing and recording data in DNA - in vivo approach Source: https://www.nature.com/articles/s41576-019-0125-3

A Promising Alternative?

Despite the advantages of utilizing DNA to store data efficiently, there are several challenges and drawbacks that we might encounter. For example, it takes a lot of time to write and retrieve data, unlike in modern data storage devices. The degradation of DNA can be prevented only if it is stored under proper conditions - low temperature and low humidity in the absence of ultraviolet rays. It is also important to develop better error-correcting mechanisms to improve the accuracy of the information obtained from the data. Hence, tremendous efforts are required to tackle these problems.

On the other hand, there is huge scope for DNA being the ideal storage device in the future due to its capacity to store enormous amounts of data in low volumes. It is fascinating to note that theoretically, all the available data present in the world can be stored in just a few hundred grams of DNA. Moreover, the ease of replication of DNA allows copying of data in constant time regardless of the size of the data. In comparison, copying times of digital storage devices are proportional to the size of the data.

In conclusion, as long as life exists on our planet, DNA can offer an eternal form of data storage that will never become obsolete!

Reference

Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. Nat Rev Genet 20, 456–466 (2019). https://doi.org/10.1038/s41576-019-0125-3

Unravelling Retroactivity: An Enigma

Roshni Shetty

Today, synthetic biology has grown leaps and bounds and transformed with applications and approaches inspired by electronics and software design. For instance, the amplified bio-sensing, timed genetic circuits, oscillators and switches. Modules in a biological system can be thought of as a network of reactions. Our goal is to modularize, device standardized BioBricks such as coding sequences, promoters, ribosomal binding sites and terminators to easily design, assemble pathways and engineer biological circuits to perform novel tasks, such as to achieve a predictable output behaviour with a defined input.

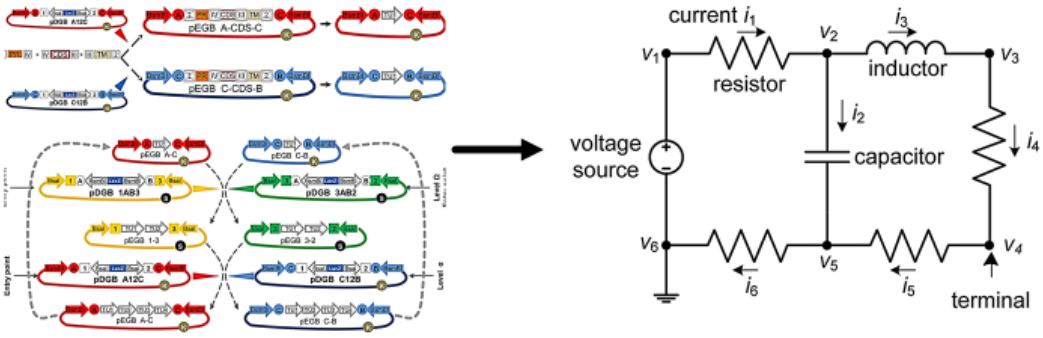


Image Source: https://journals.plos.org/plosone/article/figure/image?size=large&id=info:doi/10.1371/journal.pone.0021622.g003
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0049126

Just like in analogue and hydraulic systems, impedance or load like effects can arise in a biological system on the interconnection of components. This phenomenon is called 'retroactivity'. When the modules are interconnected, retroactivity is characterized by signals from downstream modules (that receive the arriving signals) to upstream modules (that send the original signals).

Nonetheless, an electrical engineer's life is not as complicated as a systems biologist! An electrical engineer does not have to worry about the components changing dynamically, for example, the dimensions of capacitors and resistors are almost constant, and lines of code do not rewrite themselves, but in a genetic circuit, random mutations can change everything! Besides, the solid-state transistors can be explained using simple Boolean logic, whereas computing the functionality of transcription factors using Boolean logic is a challenging task. Furthermore, electric circuits are modular, hard-wired, and the energy-loss and efficiency well-estimated, but a biological circuit operates in a diffusive environment and the molecular cross-talk cannot be ignored. It is also difficult to measure the leaky expression of the system. Moreover, an electrical engineer commonly deals with nano-second transitions in logic circuits, whereas a systems biologist needs to study gene-regulatory circuits that may operate on minutes to hours' scale.

In other words, a systems biologist has to deal with all the perplexity and complexity!

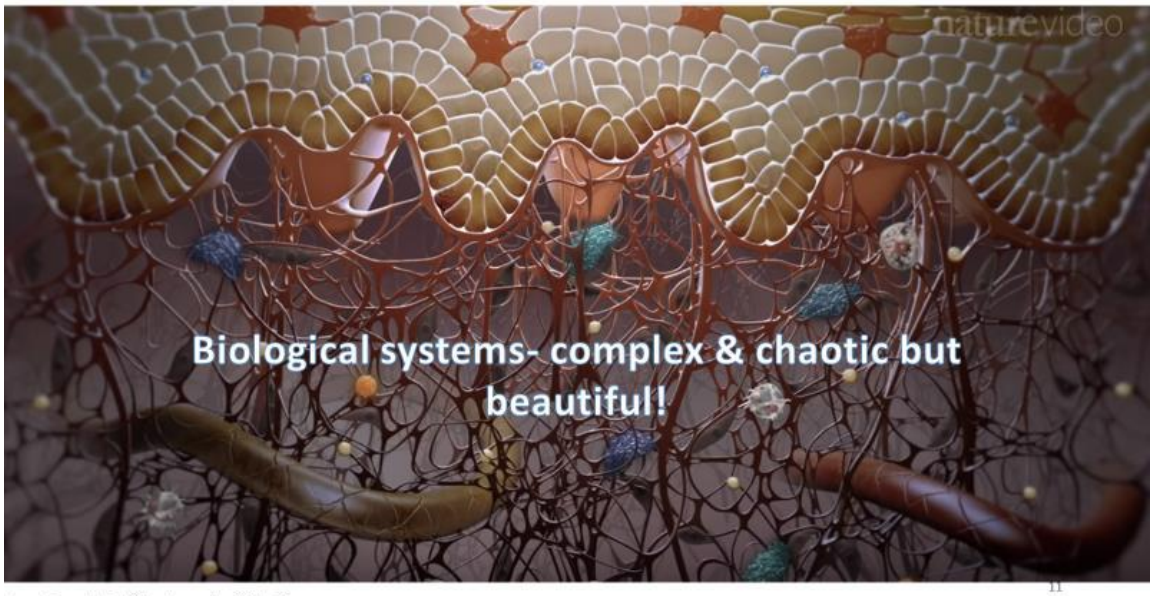


Image Source: NatureVideos-Immunology in the skin

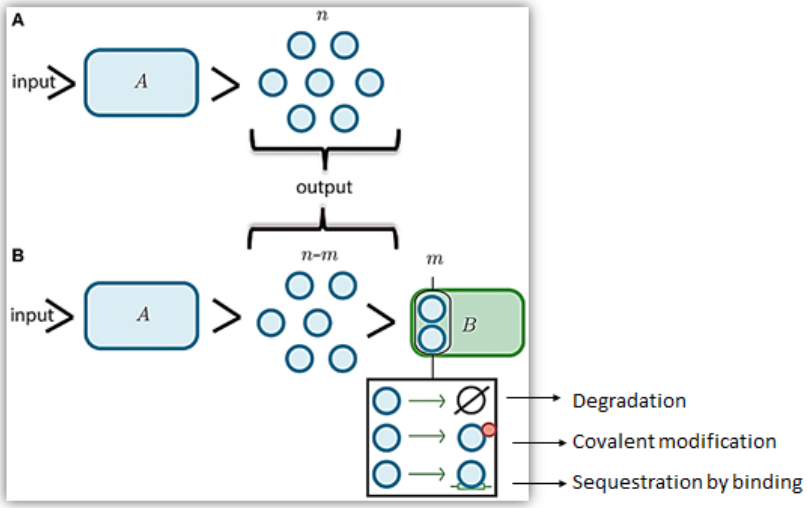


Image source: https://www.frontiersin.org/articles/10.3389/fbioe.2015.00085/full#B58

Let us understand retro-activity with the above diagram.

The function of module A is to produce a finite set of ‘n’ molecules when an input signal is given. However, a decrease in the output of module A by 'm' molecules is observed when connected to second module B. These ‘m’ molecules could have been degraded, covalently modified or sequestered thereby decreasing the output of A.

Therefore, it is evident that retroactivity is a property of both natural and synthetic systems and depending on the strength, the downstream molecules can significantly change or even disrupt the functionality of the upstream molecules.

To enhance our understanding of retroactivity, it is essential to characterise modules. **What is a module?** As per the dictionary, a module is ‘an independent unit that can be used to construct more complex structures.’

In a circuit, it becomes essential to characterize the functional modules. These are composed of different kinds of interacting molecules and perform specific, discrete functions that cannot be easily predicted from studying the components in isolation. Hartwell (1999) defines a functional module as “features that act together in performing some discrete physiological function are semi-autonomous in relation to other functional modules”.

The next apparent question that arises is - **Are biological circuits truly modular?** We can say that at a fundamental level, there is a modular composition in biological systems consisting of genes, operons and horizontal transfer elements. But, the constant inter-linkage between parts (such as covalent modification of enzymes and transcription factors), participation in multiple pathways and intertwining of functions makes it a difficult task to establish the structural and functional boundaries that delimit modules.

Thus, in a biological system, there is a balance between modularity and connectivity. Hence, we can conclude that 'biological systems are virtual units, in other words, QUASI-MODULAR!'

Structures, molecules and complexes are shared across different pathways. The limitations that lead to autonomy are highly dynamic and can change the structure at any given time. How biochemical modules can be demarcated is still an unanswered question and is being studied. Some approaches exist based on network-clustering methods applied to experimental data (Rives & Galitski, 2003), and for developing a theoretical framework for the analysis of modular networks (Bruggeman, Westerhoff, Hoek, & Kholodenko, 2002). Three biologically motivating criteria for defining functional units are: (1) common physiological task, (2) common genetic units and (3) common signal transduction network. Saez-Rodriguez et al. (2005) introduced retroactivity in biological systems and proposed that modules in cell signal transduction networks can be established based on minimal retroactivity interactions.

Why is it important to consider retroactivity? It is crucial for a deeper understanding of the interconnection between the functional modules and their impact on the overall behaviour of the cell. This influences the need for novel experimental techniques and theoretical frameworks to formally characterize and quantify effects of retroactivity. Further, retroactivity also provides an opportunity to utilize mathematical and computational models. These may unravel the properties of previously studied motifs and patterns of interconnection; the capacity to minimize or increase retroactivity.

How does retroactivity manifest in a living cell in the first place? Transcription factors may drive too many target promoters, or nucleosomes could restrict the binding of a transcription factor to its target. Due to squelching and enzyme degradation, large amounts of regulatory elements may deplete the cell's resources, leading to a reduction of the system's performance. Further, there may be "non- functional" transcription factor-binding sites which can act as loads. The Cascades of covalent modification cycles, methylation–demethylation, activation–inactivation of GTP-binding proteins, phosphorylation– dephosphorylation manifest in several regulations. So far, it seems as though retroactivity is bad and is reducing the performance of the system. But in certainty, retroactivity can have important functional roles in living systems!

Retroactivity plays an important role in integrating information into the central nervous system from different senses and dictates the organism's behaviour. The generation of action potential and regulation of neurotransmitters in nerve cells also display retroactivity. Retroactivity can additionally explain the evolution of certain proteins such as histones, actin and tubulin, that interact with many others and have altered very little over time. This can be due to the very reason for their participation in many different interactions. The decision-making module in bacteriophage lambda from lytic to lysogenic cycles is also a classic example of retroactivity. Likewise, the regulation of our fantastic immune system and the cross-talk between its various players is an excellent example!

Finally, let us review the pros and cons of retroactivity meant for a system in synthetic biology. Retroactivity can cause delays and disrupt systems. While designing a gene circuit, the goal is to ensure that it is robust to perturbations imposed by constraints on its design. However, only a small fraction of the possible circuits formed by interconnection can execute a desired function. According to Alexander et al. (2009) "knowledge of network structure is often not sufficient to infer function, and dynamic modularity can exist in the absence of structural modularity." Therefore, shaping a biosynthetic system in a modular way is a complicated task.

Modularity promotes robustness to component tolerances and allows reconfiguration for new conditions and the capacity to test functions. It improves the organization of mechanisms addressing local problems in a network. However, retroactivity has a homeostatic role, which is important for living systems. Retroactivity transfers the system from an ultrasensitive response regime to a lower sensitivity regime. It is clear that retroactivity has a functional role and is not just a nuisance signal from its manifestation in living beings.

Whether biological systems tend to minimize or potentiate retroactivity is a matter of debate in the current scenario. Quasi-modularity can be a survival strategy because preserving functions under diverse settings is imperative for life. Therefore, to fully realize the potential and promise of synthetic biology, and understand the comprehensive processes we must appreciate and consider the complexity and specificity!

Reference

Pantoja-Hernández L and Martínez-García JC (2015) Retroactivity in the context of modularly structured biomolecular systems. Front. Bioeng. Biotechnol. 3:85. doi: 10.3389/fbioe.2015.00085.

Hacking Our Own Immune Cells to Fight Disease

Sagarika Nath

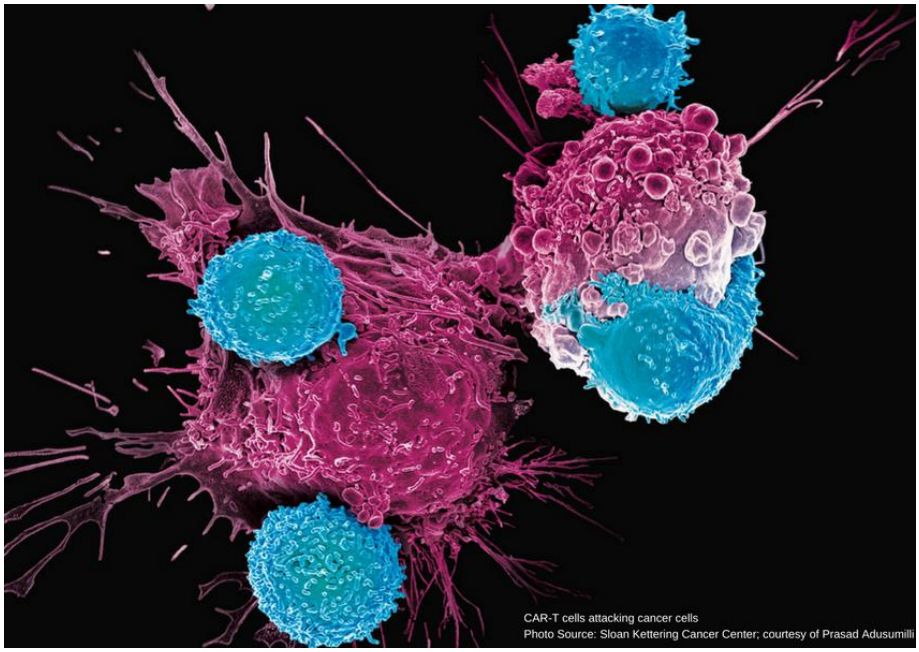
Our immune systems are what silently protect our complex bodies from the daily onslaught of millions of pathogens, injuries, -both internal and external- growing tumours and even seemingly harmless foreign particles like pollen!

Miraculously, they run without glitches most of the time, but sometimes they may make mistakes, either failing to remove cancerous cells or attacking healthy tissues in the body. The immune system is powerful and well-coordinated and therefore ideally set up to protect complex organisms from pathological insult to the fullest extent, but when it does fail, it can be fatal. This is why biologists, with the growing array of tools available to them, are keen on "hacking" immune cells and tweaking what's already within us slightly to better perform a wide variety of tasks. They could become our tiny soldiers (scientists posing as their remote commanders) who find and destroy tumour cells, pathogens, subdue painful inflammation, and self-destruct when required.

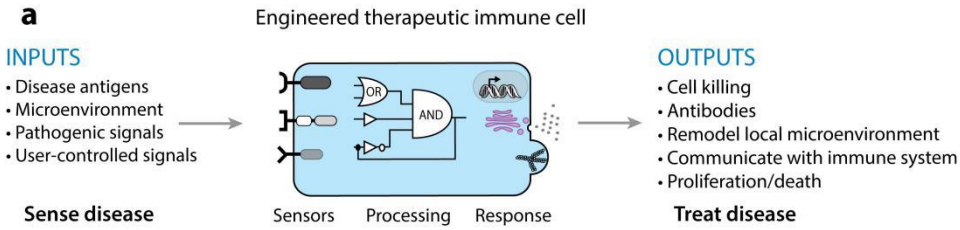
While we are a long way off from that level of reliable and all-encompassing control, the idea that we can use our immune systems to our advantage in therapeutics has gained a lot of momentum in the last few decades. Most notably, a type of immunotherapy that uses engineered antibodies blocks the 'immunosuppressors' (chemicals that block our immune cells from acting) that function around tumour cells and prevent our little soldiers from clearing them out. Why is this important? This kind of cancer immunotherapy has been approved and been saving lives from cancer for many decades, winning the original scientists (Allison and Honjo) behind it a Nobel Prize in 2018. This was the first Nobel Prize ever given to a cancer therapy! Since then, many other advances have started transforming the seemingly intractable worlds of our own bodies' immunities into another arena for scientists to *synthetically* make their own. Among these was the invention of the much-celebrated CAR T cell, a synthetic version of the white blood cells called T cells, known as the master regulators of the immune system. While this currently has quite a few side-effects and is ineffective against 'solid tumours' like colorectal and breast cancer, it has shown promise against blood cancers. Slowly and steadily, scientists are moving towards their goals of treating cancers as well as autoimmune diseases (like diabetes and MS), injuries and other diseases using immunotherapy.

"Wait, what's wrong with the current, familiar modes of medicine?" you may ask. While small molecules (drugs) and biopharmaceuticals are simple, much easier to manufacture and hence less expensive; immune cells are more adaptable, versatile and smarter. They're used to combinatorial inputs from within and outside the cells, and can affect complex pathways and systems and control them as well.

Molecular drugs are tools made for a single job, whereas cells, much like computers, can be programmed to employ the right group of tools depending on the situation. So the idea is that they will *sense* a wide variety of *inputs* (pathogens, injury, developmental signals, signals we give them, and abnormal physiology), *process* this information after moving across the body and surveilling tissues, to conditionally proliferate and 'memorise' the inputs, and finally give an equally wide variety of *outputs*, including communicating to other cells through chemokines and cytokines, killing infected/abnormal cells in a wide variety of ways, or neutralising a toxin. Also, immunotherapy is useful in treating diseases of the brain, where T cells can enter but it is very hard for drugs to, due to the blood-brain barrier.



A CAR T cell (a synthetic version of our own T cells) attacking cancer cells. Photo courtesy: in image.



b

	Small-molecule and biologic therapeutics	Therapeutic immune cells
Advantages	• Simple • Inexpensive	• Smart, locally targeted response (nonsystemic) • Target disease-associated signals (combinatorial) • Targeting and response are independent/modular • Complex, powerful responses (e.g., killing, secretion) • Amplification by cell proliferation • Adaptability • Durability/memory
Disadvantages	• Targets single molecule • Acts only by blocking direct molecular target • Systemic action and toxicity	• Complex • Expensive

Immunotherapy in a nutshell. a) The engineered therapeutic immune cell's blueprint. b) A comparison of immunotherapy against existing therapeutics.

Our immune systems have many different kinds of players. Some of them constitute the 'innate' or non-specific branch, while some form the adaptive immune system. Many of these can be viewed as platforms for engineering immune response. The innate immune system does not change over our lifetimes, and its abilities

are passed down from from generation to generation, only evolving over time. This arm includes phagocytes (professional 'eaters' of pathogens), of which the major types are dendritic cells and macrophages. Sometimes, we may need to kill a harmful cell without engulfing it. Natural Killer cells, aptly named for their degranulate-and-kill function, then come into the picture. These are platforms for engineering in immunotherapy. However, they require a constant supply of cytokines in order to be maintained, once inside the body, hence the therapy is infeasible. Dendritic cells can be 'vaccinated' against cancer, but this therapy as well has a long way to go.

Then comes the adaptive arm. The lymphocytes- T cells and B cells- form a major part. B cells are known to mature into plasma cells and secrete different types of antibodies. T cells also serve a variety of functions in recognition of antigen after it is 'presented' to them by the aforementioned phagocytes, and then deciding how to deal with the infection as a whole in a systematic way. Therefore T cells, being coordinators of a lot of immune events, are promising platforms for immune engineering. Apart from CAR T cells, the receptors on T cell membranes that recognise presented antigens (called T cell receptors or TCRs) have also been engineered and modified, particularly to recognise cancer-specific antigens. Immune engineering employs the interesting concept of 'Gordian Knot' solutions. In the legend of Alexander the Great, he untied the impossibly tangled Gordian Knot by simply cutting it with his sword. In the same vein, synthetic biologists are constantly working to bypass some complex processes in the immune system. For example, CAR T cells bypass antigen presentation by possessing the ability to process the unbound antigen themselves. Orthogonal synNotch receptors are another example, although a bit contrived for the current discussion.

Before blood donations, we need to match a donor's blood type to the receptor's, except in special cases. This is because each blood group- A, B, AB, etc., refers to the type of glycoproteins present on the surface of the RBCs inside the patient. If glycoproteins other than the types their bodies are used to (and recognise as 'self') are detected by our immune systems, white blood cells react and mount an attack, leading to problems like fever, urination of blood, etc. In the same way, if immune cells other than the ones our body recognises (through the interaction of a series of markers on their membranes) are administered into a patient, there can be severe complications. Hence immunotherapy currently requires the cells to be first taken from the patient, engineered, and then injected back in. This contributes to the high costs: upto USD 300,000 for cancer immunotherapy! A solution would be to bypass the recognition of self for these cells, so that 'allogeneic' cells (from another member of the same species) could be produced en masse, reducing costs. "People are working towards the possibility of more off-the-shelf therapeutic immune cells that could come from a universal donor. But we need to figure out a reliable way to modify the donor cells so that they are not rejected by the patient's immune system", W. Lim, one of the authors of the review paper, said in an interview.

In conclusion, synthetic immunology has made many strides in immunotherapy, but it has its work cut out for it and a long way to go.

Reference

Synthetic Immunology: Hacking Immune Cells to Expand their Therapeutic Capabilities" by Kole T. Roybal and Wendell A. Lim, *Annual Reviews of Immunology*, April 2017. doi: 10.1146/annurev-immunol-051116-052302

DNA Sequencing at 40: Past, Present and Future

Sahana Gangadharan

"DNA is like a computer program but far, far more advanced than any software ever created." - Bill Gates

Today, DNA Sequencing has become one of the most commonly used techniques in the field of molecular biology. It has become an integral part of various experiments and procedures since its origin. DNA Sequencing has come a long way in the past 40 years, and it currently addresses a breadth of problems for which it has proven very effective. To understand its history and applications better, it is rather important to learn about the origin of the biopolymer - DNA.

Discovery of Double-stranded DNA

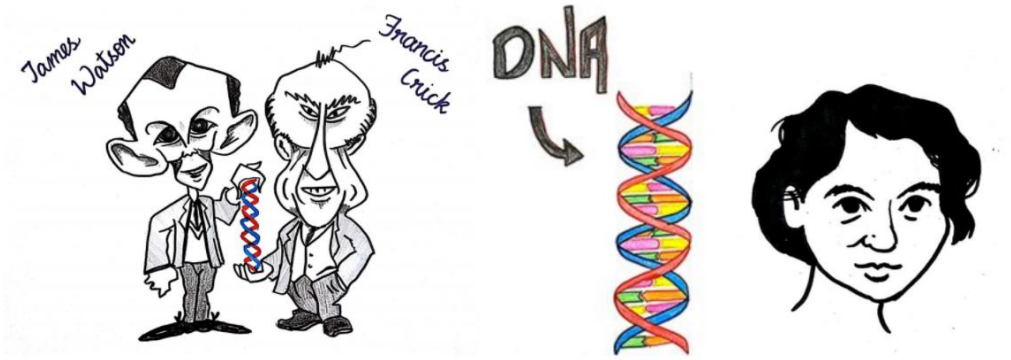


Figure 1: (Left to Right) A caricature of James Watson and Francis Crick holding the DNA, the double-stranded DNA, a silhouette of Rosalind Franklin

In the early 1950s, Rosalind Franklin performed X-ray crystallography experiments to capture the structure of DNA. Franklin's photographs were described as "the most beautiful X-ray photographs of any substance ever taken" by J.D.Bernal. Around the same time, James Watson and Francis Crick put together available information from various sources, which included X-ray images and model building techniques, and solved the most central question in molecular biology that had baffled other scientists for decades.

But when did Sequencing start?

Fred Sanger, Gilbert and Maxam built the initial blocks of DNA Sequencing. Sanger established the sequence of insulin, making it the first protein ever to be sequenced, followed by Gilbert and Maxam who deciphered the sequence of the lactose-repressor binding site. However, sequencing DNA was still a difficult task. Both these parties then individually came up with different methods of DNA Sequencing for which they were given the Nobel prize in 1980. Sanger's protocol involved four extensions of a DNA primer which involved a chain-terminating nucleotide each. On the other hand, in Gilbert's method for sequencing, four separate chemical reactions were set up to create base-specific partial cleavage. In both the above methods, we get to know the relative positions of nucleotides in the strand, thereby deciphering the entire sequence. Then came another technique, called Shotgun Sequencing, which involved chopping the DNA strand into multiple pieces, sequencing each strand individually, and then aligning them based on their overlap.

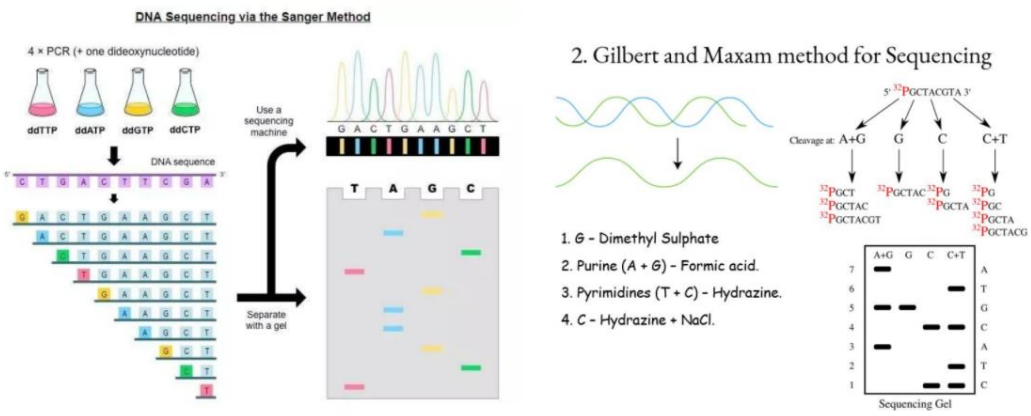


Figure 2: (Left to Right) Sanger Sequencing protocol, Gilbert and Maxam method for Sequencing Reference - https://www.onlinebiologynotes.com/sangers-method-gene-sequencing/

Ever since these mechanisms came into existence, the amount of sequenced data generated grew exponentially! Hence, additional tools and central data repositories such as GenBank-DDBJ, NCBI, etc. were set up. BLAST, a search tool which allows one to align their DNA strand with a previously annotated one, has now proved extremely beneficial to the entire scientific community. Sanger sequencing machines, which use bots to perform the experiments for us, were developed later. How amazing, right?

Sequencing the entire human genome? Possible?

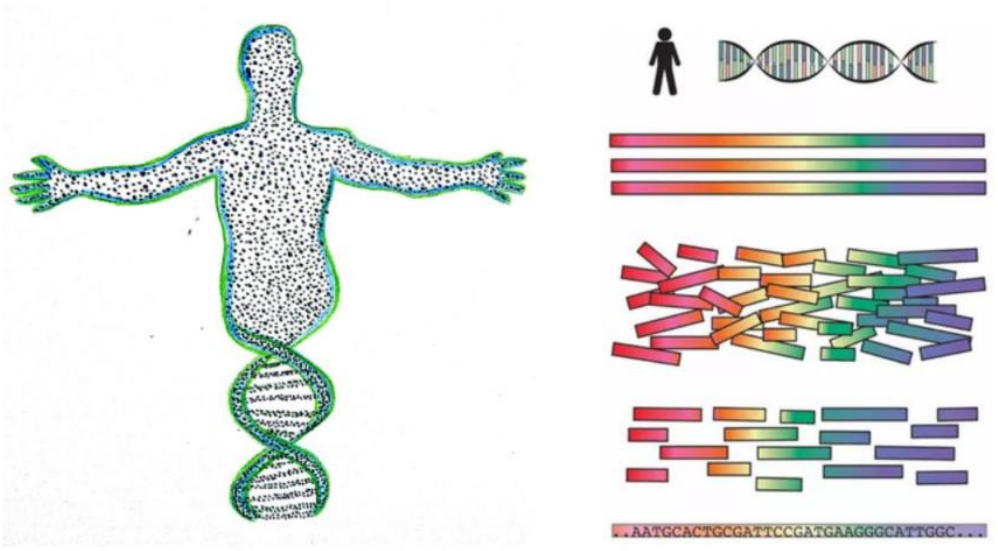


Figure 3: (Left to Right) A figurative model depicting synthesis of Human Genome, Shotgun sequencing applied for the HGP. Reference - http://sitn.hms.harvard.edu/flash/2019/lessons-from-the-human-genome-project/

With all the available data and documentation, scientists next set out to sequence the entire human genome. This project was fragmented into multiple sub-projects, and each was assigned to a different country. The Human Genome Project (HGP), one of the most ambitious scientific projects ever undertaken, took 13 years to sequence the entire human genome, thereby accomplishing a monumental goal at the end of it. But wait! Every task needs competition. Doesn't it? In parallel with the HGP, Celera promised to sequence the genome effectively and efficiently, but in three years. At the end of the game, the results were analysed, and it was concluded that both the HGP and Celera had similar performance rates. Hence, the competition was tied.

Efforts on sequencing thenceforth were reduced by a large fraction, as the above techniques and the HGP laid the groundwork for thousands of scientific studies associating genomes and other problems.

The Next-Generation is always smarter!

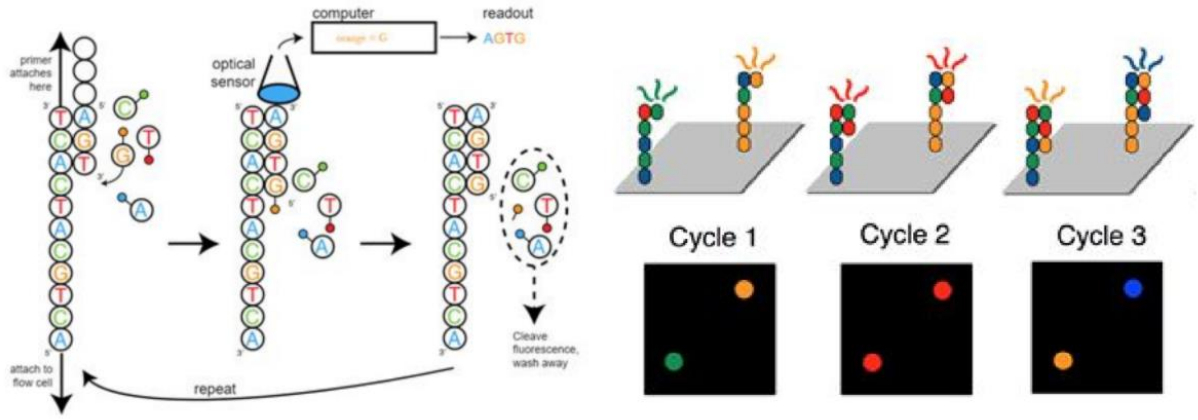


Figure 4: (Left to Right) Depiction of Sequence by Synthesis, (right-top) depictive model of how DNA is bound to the flow cell, (right-bottom) fluorescence observed on the flow cell. Reference - https://binf.snipcademy.com/lessons/ngs-techniques/illumina-solexa

The invention of *'Next-generation DNA Sequencing'* (NGS) marked a new era for the field of genomic analysis and documentation. NGS technologies provide higher throughput data for lower cost and enable population-scale genome research. Instead of carrying out one reaction in a tube, the DNA fragments are immobilised on a flow cell where the templates have common access to one reagent volume. While the DNA molecules are being synthesised, polymerase-mediated incorporation of fluorescently labelled nucleotides occurs, with each nucleotide being tagged with a different fluorescent dye. With the addition of nucleotides, the tags fluoresce which is captured by the machine, thereby enabling the sequencing of the DNA strand. This fascinating technique is also known as 'Sequence by Synthesis' (SBS) and is prevalent in all modern-day sequencing methods.

Fancier than NGS? You got to be kidding!

Around the time when NGS was ruling genomic studies, there was yet another revolutionary technique that came up. It had even higher throughput and was able to sequence larger strands of DNA within a few minutes! The first of this kind is called the PacBio - *Single-Molecule Real-Time Sequencing technique*, in which time and space complexity is way lower than that of any of the previous methods that we have reviewed. The DNA Polymerase enzyme is held in a fixed position, and the DNA strand moves around it and is sequenced while being synthesised. Now, if you had thought that any of the above techniques are groundbreaking and exhilarating, then wait! The Lord of the sequencers is yet to be revealed. The second of this kind is the most recent one - *Nanopore Sequencing*. Here's how it works. A single-stranded DNA fragment is sent through a nanopore and, based on the electrostatic interaction of nucleotides with the pore, the sequencer records the plausible nucleotide that was

encountered. The machine looks like a regular USB/pen-drive, weighs only 70 grams, and can sequence the entire human genome in less than a day!

How cool!

Applications of Sequencing

For its first 25 years, the primary purpose of DNA Sequencing was the partial/complete sequencing of genomes. However, with advancements in the field, the range and scope of DNAseq applications go from Plank to Parsec.

Today, DNA sequencing is applied in varied key areas marking itself as one of the most popularly used techniques in the fields of synthetic and molecular biology. *De novo* genome assembly requires DNAseq on a fundamental basis so that unknown stretches of DNA could either be sequenced first or matched with an already annotated sequence. Sometimes, when we lack confidence in previously available sequences, or in case we want to study the evolutionary perspectives of a gene, genome resequencing proves to be highly useful. It helps one identify mutations in particular fragments, or even reduce the error rates of the available sequence by resequencing and comparing with the ones in the repositories. Yet another important application of DNAseq is observed in clinical practice. Medications, diagnostics and treatments could be much improved if the sequence of the defective gene is available in certain diseases such as cancer.

What future awaits us?

"There are millions of species on earth (and far more extinct species), each with a genome waiting to be sequenced, as well as countless microbiomes and metagenomes." With such a genetic diversity amongst species, it is rather important to analyse our pattern of evolution by sequencing each one's genome and performing comparative studies. The nanopore sequencer will act as a real-time portable sensor which can analyse the sequence of the genome just like there are devices to analyse the pH or temperature of any solution. Other unconventional applications of DNAseq are to use DNA as a storage device by treating it as an alternate way of representing information. This, again, would require DNA sequencing.

In conclusion, DNA sequencing is essential to a molecular biologist, just like how a microbiologist would perceive a microscope. The future certainly holds many more improvements and innovations with surprising and contemporary applications of this technique. So, stay tuned!

Reference

Shendure, J., Balasubramanian, S., Church, G. et al. DNA sequencing at 40: past, present and future. Nature 550, 345–353 (2017). https://doi.org/10.1038/nature24286

The new frontier with genome engineering with CRISPR-Cas9

Sai Guha C.

Doudna and Charpentier were the principal investigators in the labs that published their paper detailing CRISPR's adaptability to gene editing[2] in Science way back in 2012. Their work drove the way for a new frontier in genome engineering, revolutionising our perspective on gene editing once and for all. It is only befitting that we take a look at the way they review the impact that CRISPR-Cas9 had on the field.

Gene editing is not a new and certainly not a novel concept. It has been around since the late 70's, when gene replacement was first achieved in yeast. There were many discoveries along the way, including and not limited to the discovery of Zinc Finger Proteins, the discovery of Homology Directed Repair and non-homologous end joining, all new methods and techniques that slightly furthered the case that we might be able to, one day, edit the human genome. Then, in the 1990's, scientists across the world managed to, for the first time, sequence the entire human genomes. Everyone thought that this would change the way we deal with human genetics forever. But none of these discoveries had the impact they'd been advertised to have, until now, that is.

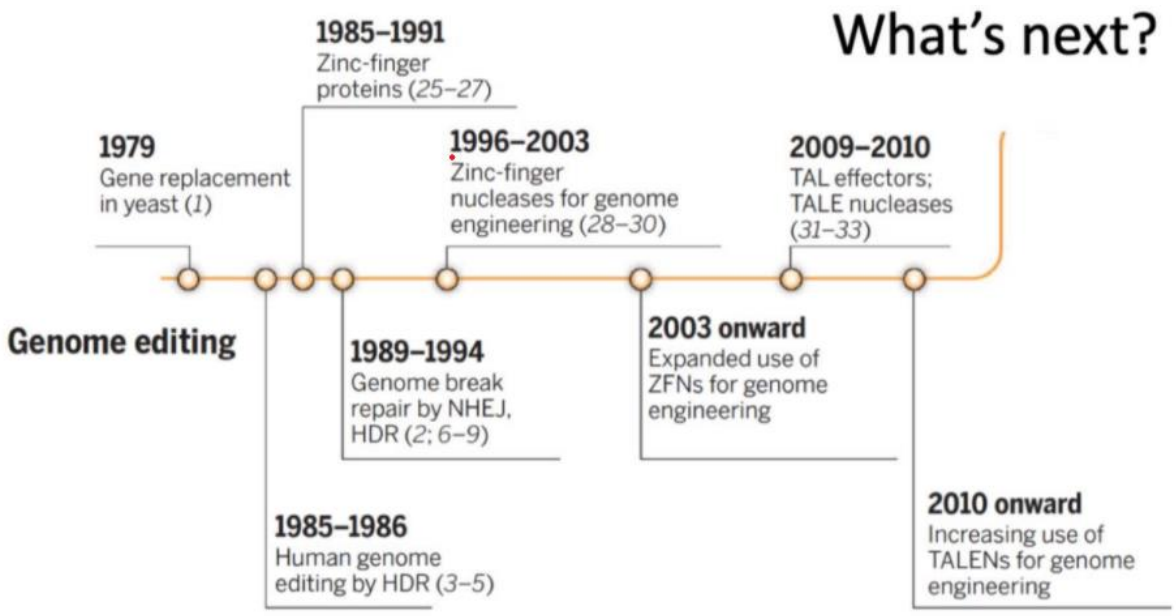


Fig 1. The timeline of progress on genome editing (without CRISPR).1

History of CRISPR-Cas

In the 2000's, a range of laboratories began investigating CRISPR for unrelated reasons. They were found in certain bacteria and archaea, which seemed to be clustered repeats of certain genetic sequences. It was later discovered that the sequences seemed viral in origin or resembled certain plasmids. Further down the line, it was proposed that CRISPR-Cas was an adaptive immune response to past invasions, and that the sequences were antisense reminders of invading sequences[3]. The same was established to be true in *Streptococcus thermophilus*. After this was it first hypothesised that bacteria could be weaponized to act as a deterrent to

phages. This was tested successfully in 2008, and CRISPR RNA's were shown to act as guides to locations that had to be spliced from invading genomes.

CRISPR-Cas: How does it work?

The operation of the CRISPR-Cas system is simple. Let us look into some basics on DNA first. DNA is a polymer comprised of four molecular bases, commonly represented as A, G, T and C. The structure of DNA is based on base-pairing, where A and T pair with each other and G and C pair do likewise. It is this pairing that results in a double stranded structure of DNA, where each strand perfectly binds with the opposing strand[4]. RNA has a structure that is similar to DNA, except for the T being replaced with a U. RNA is *transcribed* from DNA, and the RNA obtained is almost identical to the primary or the *Watson* strand of DNA, named after one of the researchers who discovered the structure of DNA. When the RNA is exported out of the nucleus, it is translated into proteins, which drive functions around the body. RNA, in and of itself, has the ability to base pair with a strand of DNA. RNA exists in nature usually as a single stranded molecule.

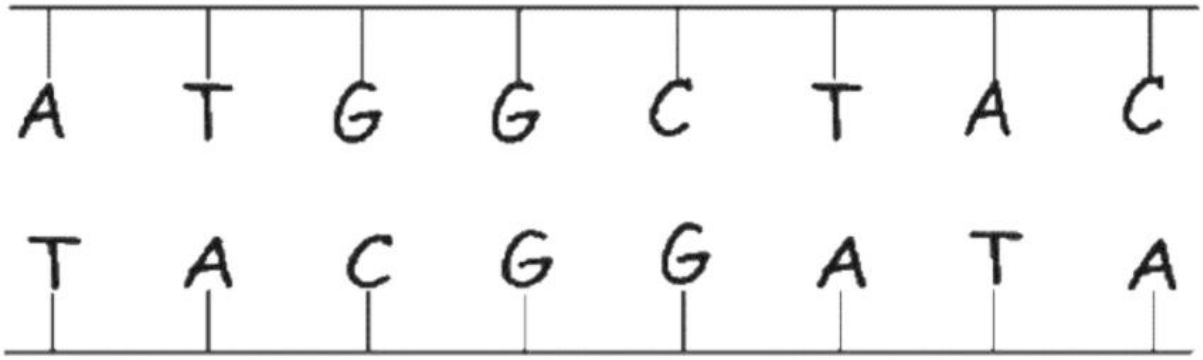


Fig. 2. DNA Base binding and complementarity.

In natural systems, invading virus/plasmid's RNA have been *reverse transcribed* or reversed into DNA and subsequently integrated into host genomes in the form of repeated sequences. Hence, CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. All such sequences are clustered in a region in the host genome called the CRISPR locus. The Cas protein, prepared by the cellular machinery, and the RNA similar to the viral RNA, that was prepared by the host, form a complex and begin surveillance of the cytoplasm. This is known as the CRISPR-Cas surveillance complex.

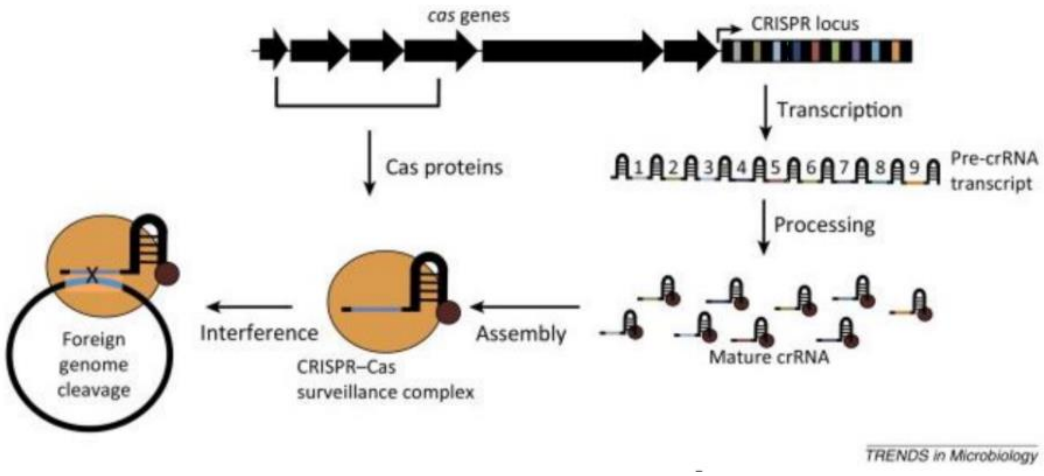


Figure 3. CRISPR Cas in Nature.[5]

The surveillance complex, now containing the CRISPR RNA's, surveils the cell and uses the RNA to home to the sequence with which it shares complementarity. Once it homes in, the CRISPR-Cas system scans for a motif called a PAM motif, which must be present at the target of interest. If the PAM motif is found, the CRISPR-Cas complex binds to the DNA and induces what's called a double stranded break (DSB). Once the break has been created, the cellular machinery of the invading pathogen stitches the broken area up, but without the sequence that had been cut. If the sequence coded for a protein that increased the pathogenicity of the invader, the invader is now *disabled*.

CRISPR-Cas: What can scientists do with it?

This very process can be exploited by scientists to edit the gene of their choice. The CRISPR RNA is around 23-55 nucleotides, and homes in to a sequence with perfect complementarity to induce a DSB. If the guide CRISPR RNA is engineered to reach the gene of interest, a DSB can be artificially induced anywhere. Techniques such as homology directed repair (HDR) also allow the cell to introduce a new sequence of interest at the location of a DSB. So, to modify sequence A into sequence B, the CRISPR-Cas system must be introduced into a cell, with the RNA sharing complementarity with A. Once the break is introduced, a HDR is initiated with sequence B, to make sequence B take A's place in the genome. If the intent is to simply delete a gene, sometimes referred to as a *knock-out*, an alternate pathway referred to as non- homologous end joining is utilised. However, it is important to note that the sequences that are cut are usually very small, and the process must be tweaked to remove entire genes.

There are other ways CRISPR-Cas can be exploited. Since they recognise highly specific regions on the genome, they can also be used to map the structure of chromosomes in the relaxed state. They can also be used to make *epigenetic* modifications, modifications made directly to the genome by adding a structure to inhibit/ induce the gene.

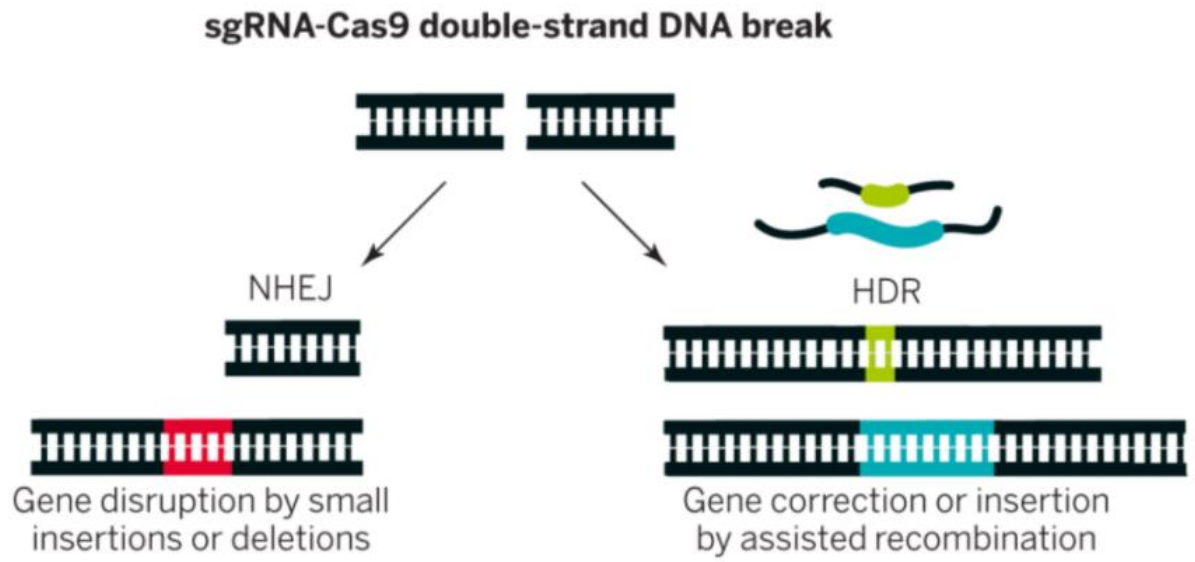


Fig. 4: What to do after a DSB is induced.[1]

The advantage that CRISPR has over all its predecessors is specificity and accuracy. The Cas9 protein will not bind to any region that doesn't have both perfect complementarity to the RNA sequence it holds and the PAM motif mentioned above. While the latter might seem like a disadvantage, scientists seem to be on the verge of developing workarounds to that limitation.

Where do we go from here?

CRISPR-Cas9 has been revolutionary in academia and the impact has already started to extend to the clinical world. Human trials to treat genetic disorders are underway, with one woman in the United States being treated of sickle-cell anaemia using CRISPR therapy. While clinical applications progress, CRISPR has also given researchers vast potential to understand complex genes across various organisms and a quasi-regulatory mechanism, which can be exploited to understand the inner workings of various organisms. CRISPR is, hence, an invaluable tool in the toolkit of the microbiologist, the molecular biologist and the geneticist and most anyone trying to understand life.

But as we go down this path utilising one of the most powerful tools ever innovated, it is also the pertinent that we have discussions in the public forum on the ethics of different kinds of gene editing, and how legislation would be framed in this regard. A review of the legality worldwide, conducted in 2016[6], characterised large portions of genome editing policy worldwide as 'vague' and not sufficiently comprehensive. To truly utilise these amazing tools, we must make sure we account for the deleterious effects they could have when placed in the wrong hands, and ensure, to the greatest extent, that the tools are not misused as we move forward into a new frontier of human history. *(The author is a junior undergraduate student at the Department of Biotechnology, Indian Institute of Technology Madras.)*

References

Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR- Cas9. Science 346, (2014). doi: 10.1126/science.1258096.

Diagnostic Trends Using Synthetic Biology

Sankalpa Venkatraghavan

"Scientists study the world as it is, engineers create the world that never has been." - Theodore von Karmen

Introduction

Imagine a world in which we can build organisms to do our bidding. Maybe some of them are innately able to perform these tasks – like detecting iron, synthesising drugs, or producing plastics. How do we harness these capabilities? It has been well established that we can use microorganisms to produce useful chemicals. We've been producing ethanol using yeast since time immemorial. But can we do better? Can we engineer microorganisms to say, do our homework for us?

Not yet! But we've made great progress. One can draw analogies to the domain of electronics and its applications. First, we saw the emergence of parts like resistors, capacitors, and transistors and many systems like Integrated Chips (ICs), filters, oscillators, logic gates, and counters were made by combining these parts. These systems have resulted in a large number of applications ranging from mobile phones to robotic vacuum cleaners.

Synthetic biology utilises *biological* parts which can be put together to form many systems including oscillators, logic gates, and counters amongst others. This has resulted in the development of applications in many domains ranging from health care to efficient energy production. Synthetic biology straddles the world between science and engineering.

In the context of diagnostics, the theme involves the development of sensors which can be coupled to a measurable output or a visual readout. These sensors can largely be built at three scales – the whole-cell level, *in vivo sensors* and *in vitro* sensors.

Whole-Cell Biosensors

Canaries were used in coal mines to detect the release of noxious fumes. If we devised a way not to use the entire bird but rather, the smallest functional unit of it to detect fumes somehow, we would have designed a whole-cell biosensor. These can be regarded as an extension of using animals to detect changes in the environment. Traditionally, single-celled organisms were physically merged with hardware components to serve as detectors. This is a hybrid device. Modern approaches involve engineering organisms to have built-in components which can act as sensors and provide an output. Typically, the sensor is linked to a bio-luminescent output.

Fig 1: *Canary in a Coal Mine* (https://share.america.gov/english-idiom-canary-coal-mine/)

This has led to the creation of compact, portable, low-cost devices. A disadvantage is that data cannot be collected in a high-throughput fashion, but this can be mitigated by improving the companion electronics.

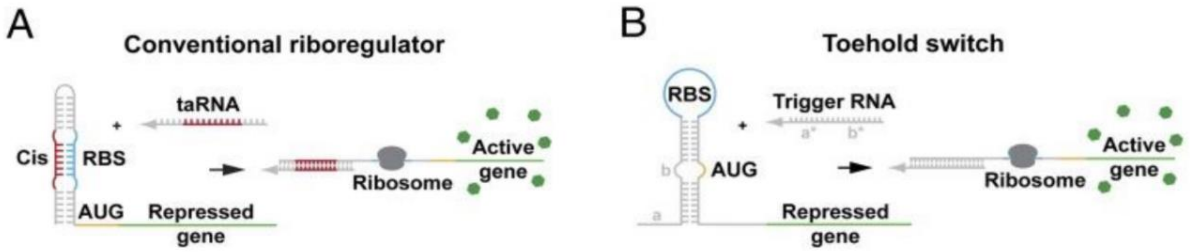


Fig 2: *A. Riboregulator, B. Toehold switches* (*Slomovic S, Pardee K, Collins JJ. Synthetic biology devices for in vitro and in vivo diagnostics. Proc Natl Acad Sci U S A. 2015;112(47):14429– 14435. doi:10.1073/pnas.1508521112)*

In a cell, a gene can be expressed only if it has a free Ribosome Binding Site (RBS) and a start codon (AUG) [Take these words to be gospel].

In Fig. 2A we see that the RBS is blocked. This can be visualised as a zipper – when it has been manufactured, the zip is bound. In the presence of an activator, the zip is effectively pulled, and the RBS is set free. When the RBS is set free, the gene is expressed. This method, however, isn't very useful in diagnostics as the zipper needs to be constructed to have the RBS, which is cumbersome if you want to use the same model to detect different things.

This idea was tweaked to develop toehold switches. Here the start codon and the RBS are blocked but, drawing from the example above, the start codon is zipped up while the RBS is just hidden. With the addition of a trigger, the zip is undone, and gene expression follows. It is much easier to construct targets to enclose the start codon (only one amino acid) which can be detected by appropriate triggers.

In the field, one can use a unique sequence belonging to a microbe to construct a complementary target so that, in the presence of the microbe, the target will be attacked by the trigger leading to the expression of some gene – say one that produces green colour. In summary, we would see green if the microbe was present.

In Vivo Sensors

While we can detect the external environment, can the same principles be used to engineer bacteria to detect the *environment* within us?

Studies have shown that it's possible to get micro-organisms to figure out what's happening inside a mammalian cell or inside a mouse. It is possible to engineer bacteria to home in on tumours in mice. They were tagged with a colour, so the colour could be tracked when UV light was shined. Synthetic circuits have been built to classify whether a cell was cancerous or not! (The classifier circuit could distinguish between HeLa and non-HeLa cell lines.)

In Vitro Sensors

So, which of these are still in the realm of science fiction? What is actually being used in practice?

Many of the techniques discussed above involve bio-safety hazards and still being developed for real-world applications. To put it crudely, they're still closer to fiction.

BUT, that's not all the field has to offer.

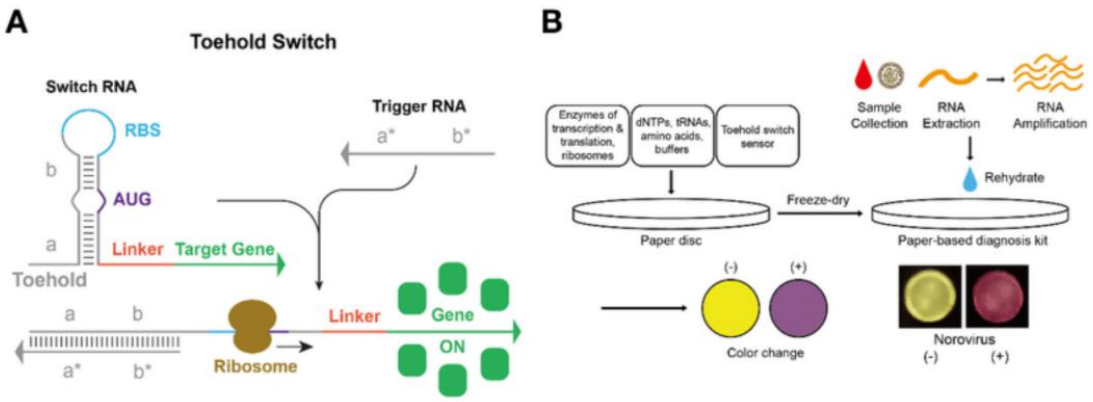


Fig 3: *A. Toehold switches, B. Paper-Based Diagnostics (Jeong, & Klocke, Melissa & Agarwal, & Kim, Taewoo & Choi, Chil-Sung & Franco. (2019). Cell-Free Synthetic Biology Platform for Engineering Synthetic Biological Circuits and Systems. Methods and Protocols. 2. 39. 10.3390/mps2020039.)*

Paper-based diagnostics have been developed very recently. These are low cost, easy to use, portable and, most importantly, very safe!

For a layperson, essentially, the paper will change colour in the presence of a microbe or a *trigger*. One could visualise pH paper, for instance.

The use of cell-free systems enabled this breakthrough application. All the proteins required to keep a cell running were isolated from a cell, freeze-dried to preserve its activity, and stored on paper. When the paper is rehydrated, the system is functional. Fig. 3 depicts how a toehold switch can be imprinted on a paper disc to be used as a paper-based diagnostics kit.

Bio-Safety

If all this talk regarding engineering organisms and using them in the real world has given you pause about its applicability in real life, fret not. These engineered organisms cannot be released into the world without appropriate testing and clearance.

It is important to keep this aspect in mind while engineering microbes or cells in the lab.

Conclusion

The application of techniques from synthetic biology for diagnostic purposes can lead to applications which are faster, cheaper and more specific. Many of these techniques are still being developed and are still in their nascent stages. Paper-based methods are particularly exciting as they can be directly deployed in the field in the case of disease outbreaks. With the development of paper-centrifuges, the *triggers* can easily be isolated and used. The safety aspects of deploying devices must always be considered, and this gap must be bridged for most of these ideas to be put into practice.

Reference

Slomovic S, Pardee K, Collins JJ. Synthetic biology devices for in vitro and in vivo diagnostics. Proc Natl Acad Sci U S A. 2015 Nov 24;112(47):14429-35. doi: 10.1073/pnas.1508521112.

Cheating Evolution: Engineering Gene Drives to Manipulate the Fate of Wild Populations

Sathvik Anantakrishnan

Introduction

Synthetic biology is a field of biology whose aim is conventionally understood to be the engineering of biological systems to serve useful purposes, and the systems in question are generally understood to be individual cells or small populations of cells. However, research in synthetic biology has yielded tools that allow the engineering of not only cells in a tube but also of entire wild populations. These tools are called gene drives.

A gene drive is a tool that allows us to 'drive' a gene of interest through a wild population, resulting in either the modification or suppression of the wild population. This is done by biasing the inheritance of particular alleles. Going against Mendelian genetics, gene drives cause some alleles to be inherited at much higher frequencies than others, thus causing these genes to quickly spread through the population. Research into gene drives has typically focused on solving two global problems: the suppression of invasive species and the control of vector-borne diseases. The latter is an active area of research as insect vectors, with their high rate of reproduction, are a model on which gene drives are easily testable and can be applied to solve critical problems. This can be done, for instance, by simply reducing the population of insect vectors or by employing more advanced approaches such as the engineering of wild mosquito populations to express genes that kill malarial parasites and dengue viruses.

Classes of Gene Drives

There are five main classes of gene drives, each with their own advantages and drawbacks. We will take a brief look at all five classes.

Homing-Based Drives

Homing-based drives involve the use of homing endonuclease genes (HEGs). The HEGs are coupled to a gene on one chromosome and recognize a specific sequence within the same gene on the other chromosome. During the formation of gametes, they disrupt the other chromosome and homology directed repair, which repairs one chromosome using the other as a template, leads to the copying of the gene drive element onto the other chromosome as well. So, all gametes will contain a copy of the gene drive element. By conventional Mendelian genetics, only half of the gametes should have a copy.

As the engineering of an endonuclease to recognize a DNA sequence of choice is difficult, the CRISPR-Cas9 system may instead be used. Here, a gRNA (guide RNA) complementary to a specific sequence as well as the Cas9 protein are expressed by the gene drive element.

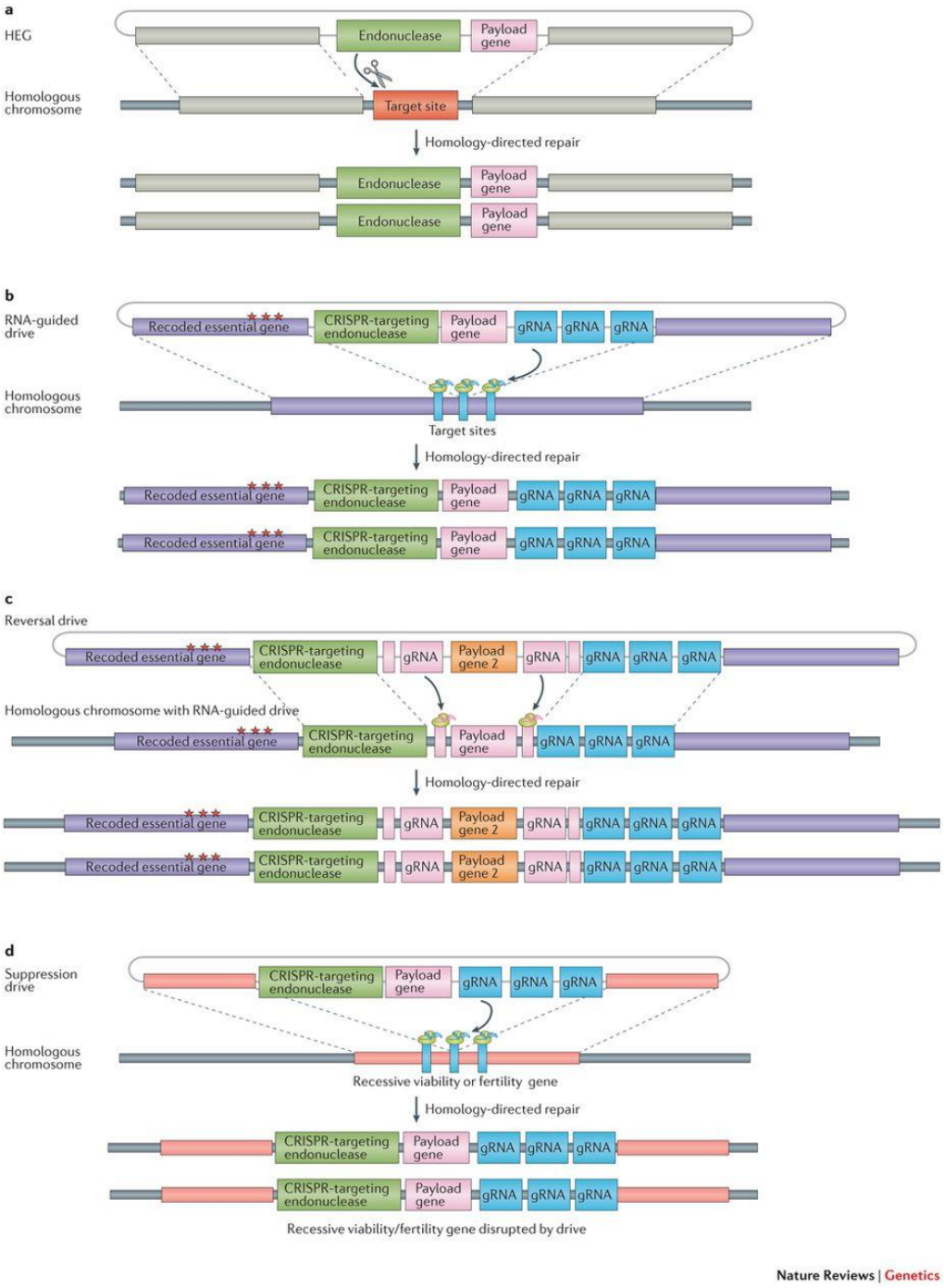


Fig. 1: HEG mechanism (a) and its CRIPR-based implementation (b), as well as the design of reversal (c) and suppression (d) drives

Homing-based drives can be modification drives when the gene drive element carries a payload gene that is to be driven through the population or suppression drives when the HEG targets an essential gene. A zygote that inherits a suppressor drive from both parents has no viable copies of the essential gene and dies.

A homing-based drive can be reversed by releasing another drive that disrupts the first gene drive. However, this does not cause the population to revert to a true wild state. So, the effects of a homing-based drive cannot be entirely reversed. It is also possible for individuals to develop resistance to such a gene drive if the disrupted chromosome is repaired without the use of the gene drive as a template (for example by non-homologous end-joining) or if the gene drive is improperly copied leading to errors in the gRNA sequence.

Sex-Linked Meiotic Drives

These drives generally prevent the formation of gametes of a particular sex that lack the gene drive element. The best-characterized drive of this class is called the X-shredder. The X-shredder is present on the Y-chromosome in males and encodes an endonuclease that cuts or 'shreds' the X-chromosome at multiple sites. This shredding takes place during gamete formation and ensures that all sperm cells produced by the male individual bear Y-chromosomes. So, males that carry the X-shredder cannot give rise to female progeny, and all their male progeny carry the X-shredder as well. This gradual rise in the male population carrying the X-shredder will eventually lead to a population collapse due to the low number of sexually-active females.

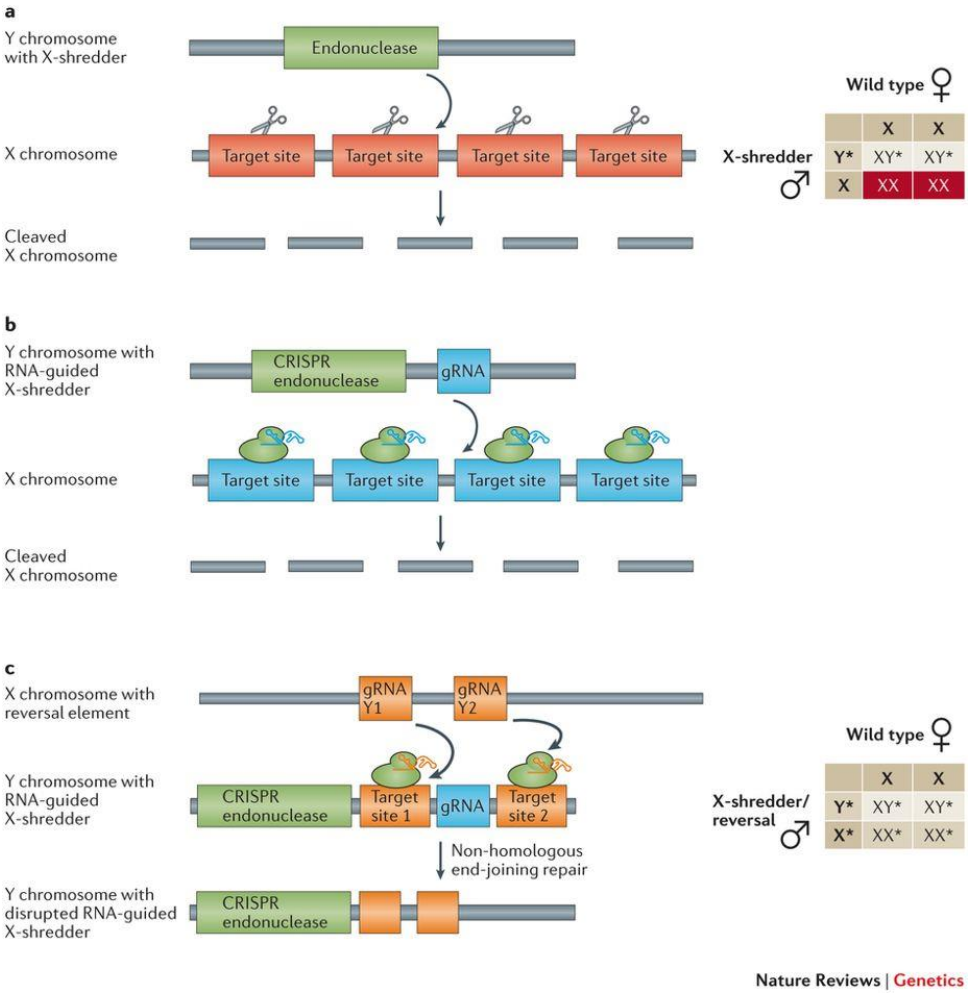


Fig. 2: The mechanism of the X-shredder and its inheritance pattern (a), its CRIPSR-based implementation (b) and a design for a reversal drive (c).

The X-shredder can also be implemented using CRISPR. Instead of endonucleases that recognize multiple sites, we can simply use gRNAs that are complementary to multiple sites along the X-chromosome.

The effects of an X-shredder can be reversed by releasing males with X-chromosomes that have gRNAs that target the gRNA of the X-shredder. This leads to the removal of the X-shredders gRNA and thus inactivates the gene drive. However, this method of reversal leaves an imprint in the DNA of the wild population in the form of the reversal gRNAs.

Medea

The maternal effect dominant embryonic arrest (*Medea*) elements encode a toxin that is deposited in all egg cells during gamete formation. The same element also encodes an antidote for the toxin. So, half of the progeny inherit the element and survive. This ensures that all offspring carry the *Medea* element. If both the male and female parents carry the *Medea* element, then three-fourths of their progeny survive.

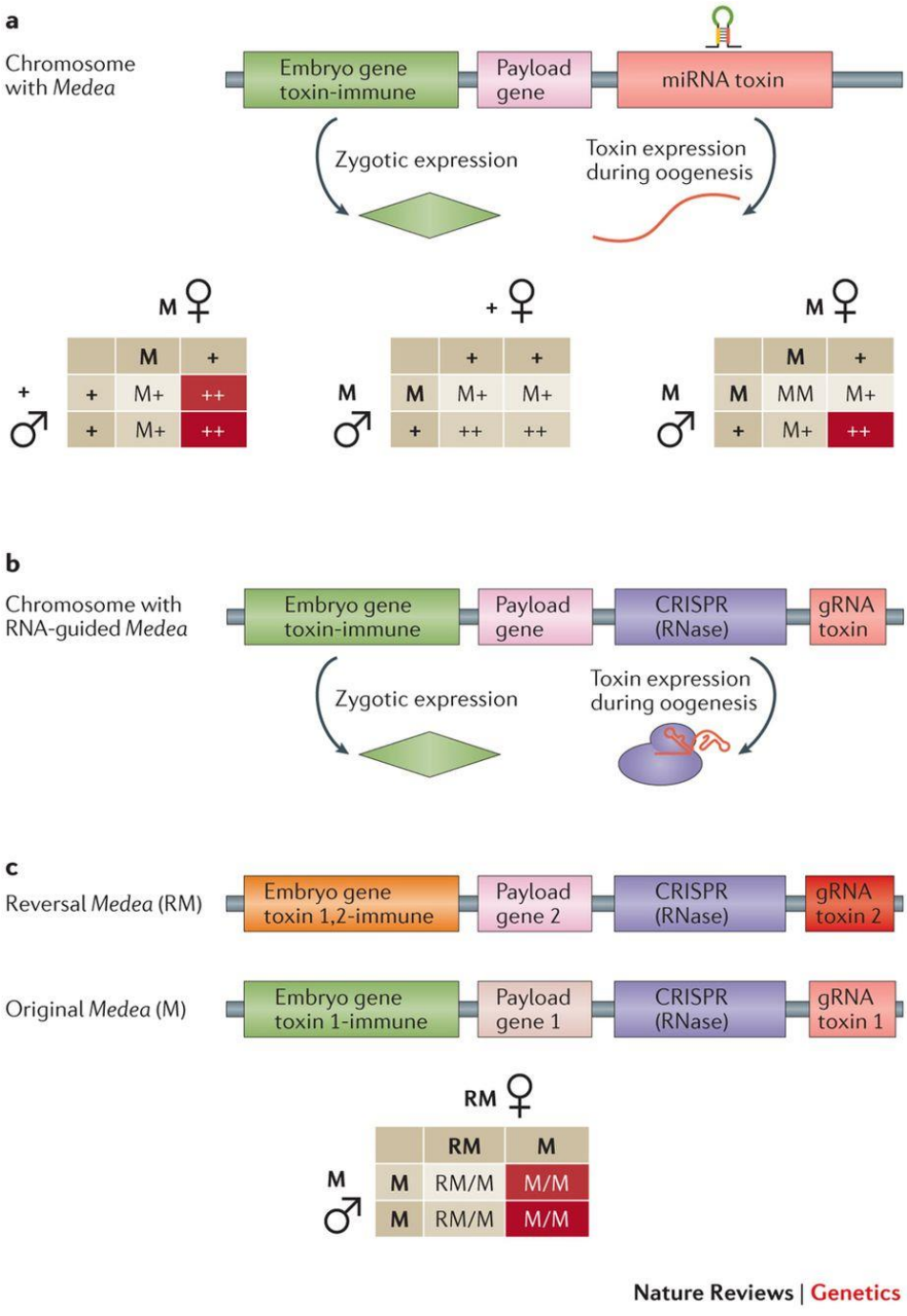


Fig. 3: The *Medea* mechanism and its inheritance pattern (a), its CRISPR-based implementation (b) and the design for a reversal drive and its inheritance pattern (c)

Medea, like other gene drive systems, can be implemented using a CRISPR system that targets an essential gene in place of the toxin and the same gene with a re-coded sequence that isn't complementary to the gRNA as the

antidote. *Medea* cannot be entirely reversed, but reversal drives can be constructed that replace the first payload gene using a second *Medea* element.

Underdominance Gene Drive

An underdominance system is on in which all homozygotes (having 2 wild-type chromosomes or two chromosomes bearing the gene drive) have greater reproductive fitness than all heterozygotes (bearing one copy each of the wild-type and gene drive chromosomes). This is achieved by having each chromosome of the transgenic individuals encode an antidote to a toxin on the other chromosome. Similar to *Medea*, this can be built using a CRISPR system and a re-coded gene in place of the toxin and antidote.

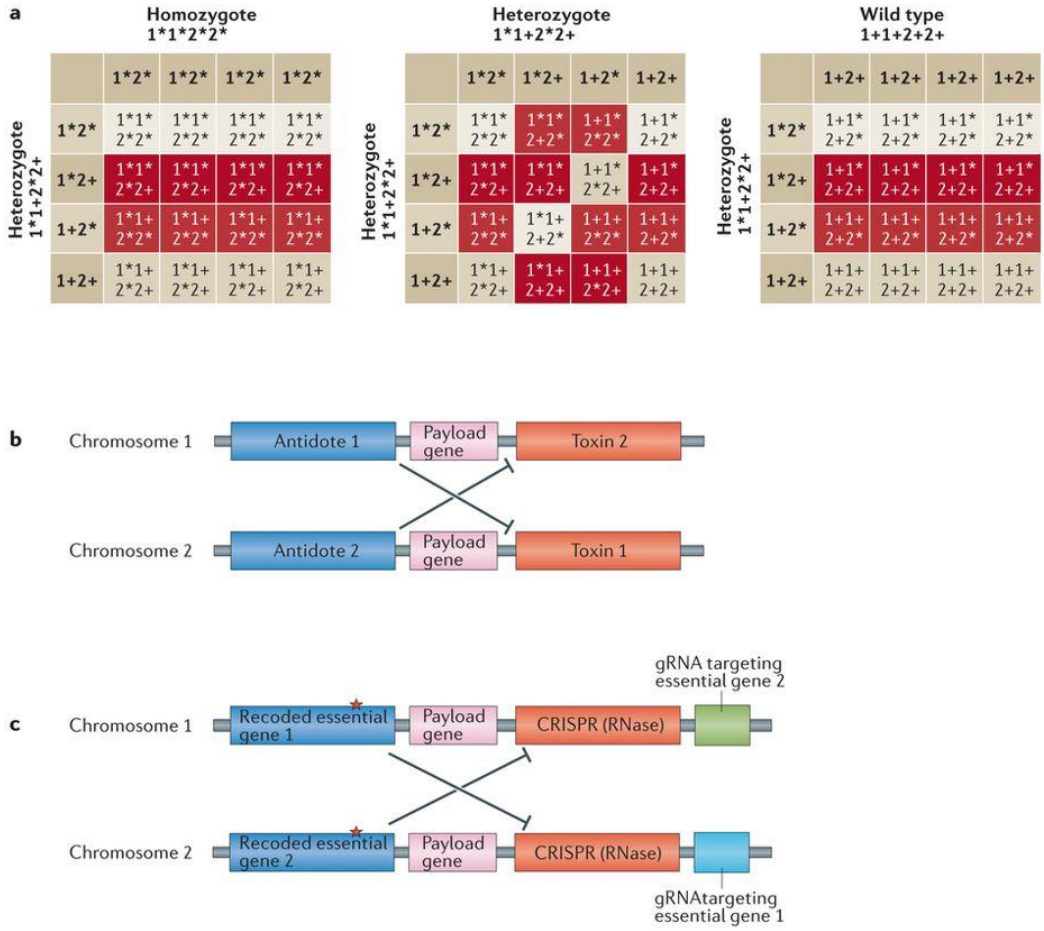


Fig. 4: The inheritance pattern of an underdominance gene drive (a), its mechanism of action (b) and its CRISPR-based implementation (c)

An interesting feature of underdominance systems is that they behave like population-level bistable switches in which the population moves towards being composed entirely of transgenic or wild-type homozygotes depending on which of the two had a greater population initially. This is because, assuming mating is random, the homozygotes with lower initial frequency are more likely to mate with homozygotes of the other kind leading to unfit offspring, while those with higher initial frequency are more likely to mate among themselves leading to fitter offspring.

This bistability implies that the effects of underdominance systems can be entirely reversed by simply releasing a large number of wild-type individuals such that the population of wild-type individuals becomes greater than that of the transgenic ones. However, this is obviously not an idea that can easily be put into practice.

Heritable Microorganisms

The last type of gene drive we'll be looking at is a remarkable bacterium called *Wolbachia*. *Wolbachia* is an intracellular parasite that has can be found in about half of all insects. It resides in a variety of different cell types including egg cells, which is how it is passed on from mothers to offspring.

Wolbachia is yet to be extensively studied and characterized. While we know very little about it, we do know of some strains of *Wolbachia* that induce parthenogenesis, male-killing, feminization and shortening of lifespan in hosts, all of which are traits that could be used to modify or suppress host populations. Much work still needs to be done before we can engineer *Wolbachia* to do our bidding, but field trials using existing strains have yielded positive results.

Concerns

The use of gene drives on wild populations could have multiple unanticipated effects, and we are not yet capable of assessing all possible consequences of the release of transgenic organisms into the wild. It is possible for gene drives to have unintended off-target effects on other parts of the host genome. It is difficult to confine gene drives geographically as individual organisms migrating out of the area may carry the gene drive element with them. There is also the possibility, however small, of the gene drive being transmitted from one species to another.

However, the damage caused by vector-borne diseases is immense. Every year, these diseases infect over a billion people and kill over a million. More than half the world's population, primarily in the developing regions of the world, is at risk of these diseases. While there are many who dismiss gene drives as being too unpredictable, perhaps we owe it to the billions of people world-over who aren't as privileged as us to continue research on systems that have the potential to alleviate untold suffering.

Reference

Champer, J., Buchman, A., & Akbari, O. S. (2016). Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nature Reviews Genetics, 17(3), 146–159. https://doi.org/10.1038/nrg.2015.34