BCG

THE BOSTON CONSULTING GROUP

Working Paper

Rewriting The Book Of Life:
A New Era in Precision Gene Editing

Elsy Boglioli and Magali Richard

September 2015

Advances in genetics - that began in the early 1960s - have led to continuous improvements in our understanding of DNA's central role in the determination of biological attributes. These advances led to numerous innovations in medicine and agriculture, such as large-scale production of insulin by bacteria for diabetic patients and plants made tolerant to herbicide to improve yield while reducing environmental footprint. This marked the entry into a new period like the first industrial revolution marked the entry into the 19th century. A second revolution began in 2000 with breakthrough advances in DNA sequencing technologies that read the information contained in DNA. This led to vast sequencing programs with the aim of sequencing the entire tree of life. It opened new areas such as personalized medicine, taking into account the genomic peculiarities of each individual.

With this easier access to DNA sequences, today we are on the verge of a third revolution that will deeply impact our lives, to the extent that computers have changed society: we are entering the era of "gene editing", following the era of "gene reading". Gene editing is the rational and precise modification of DNA sequences program in living cells and organisms. Why edit genes? For everything: from designing pathogen-resistant crops or therapeutic correction of defective genes responsible for diseases to rewriting the program of organisms to produce new sophisticated biologicals. The application possibilities are beyond imagination.

Not surprisingly, this new revolution has already sparked the enthusiasm of scientists and investors, with over $1 billion USD of venture-capital financing invested in emerging gene-editing technologies within the past two years.

Two recently developed technologies, Transcription Activator-Like Effector (TALE) nucleases and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) nucleases, make it possible to edit genes within a few weeks, as opposed to several months or years. Once a fully dedicated research subject, gene editing is becoming a routine manipulation in life science laboratories.

Combining accessibility with powerful potential, these technologies have already triggered ethical and legal debates. We face the usual switch from "Can we do it?" to "Should we do it?", similar to current and past debates on other groundbreaking innovations such as human genome sequencing today or the printing press in the fifteenth century. Still we can assume that the possibilities introduced by advances in this field will drive overall acceleration and enthusiasm rather than slowdown and reluctance.

The pace of development is indeed already accelerating, driven by the amount of money invested in these technologies. The unanswered question is what will be the next safer, more efficient and more precise gene editing technology? What is certain is that gene editing is about to change our lives in many ways.

Table of contents

1. Gene editing is revolutionizing biotechnology

1.1. From mankind's origin…

The genome can be considered as the set of instructions governing the biological functions of every living being. Biotechnology is the use of biological processes and organisms to improve human life. In that sense, biotechnology is actually a very ancient science, born 10,000 years ago from agriculture and fermentation, which has accompanied humans' quest to reshape nature to suit their needs.

For millennia, the only way to achieve that goal was laborious and time-consuming **selective breeding**, crossing plants or animals with the best attributes to generate descendants with the most optimized instruction set, i.e., the best gene characteristics.

The first significant improvement to selective breeding came in the beginning of the 20th century with the use of radiation or chemicals to induce DNA damages in order to generate new attributes. This random process, called **induced mutagenesis**, was a rather blunt method, comparable to throwing ink on a piece of paper and hoping it results in an intelligible drawing. Indeed, the DNA changes generated by induced mutagenesis are uncontrolled, randomly localized in the genome, and often harmful for the modified organism.

Despite these limitations, induced mutagenesis was used throughout the 20th century to produce a wide variety of new plant variants, many of them currently being cultivated.[1]

DNA is the physical support of the genetic information. Present in each cell and inherited from parents, it carries the **genes** that are genetic information units specifying physiological attributes such as eye color, blood type, or the propensity to develop certain types of cancer or to carry a genetic disorder. There are approximately 20,000 genes in the human genome.
DNA is composed of a sequence of **nucleotide** subunits. There are four DNA nucleotides (or **bases**): Adenine (A), Guanine (G), Cytosine (C) and Thymine (T); these are the same in every living being. The genetic information in each gene is encoded by its nucleotide sequence and this code is also universal in life (at least on earth…).

	Human	Cow	Maize	Bacteria
Genome size (10^6 base pairs)	3,400	3,000	5,000	1–5

[1] Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture.

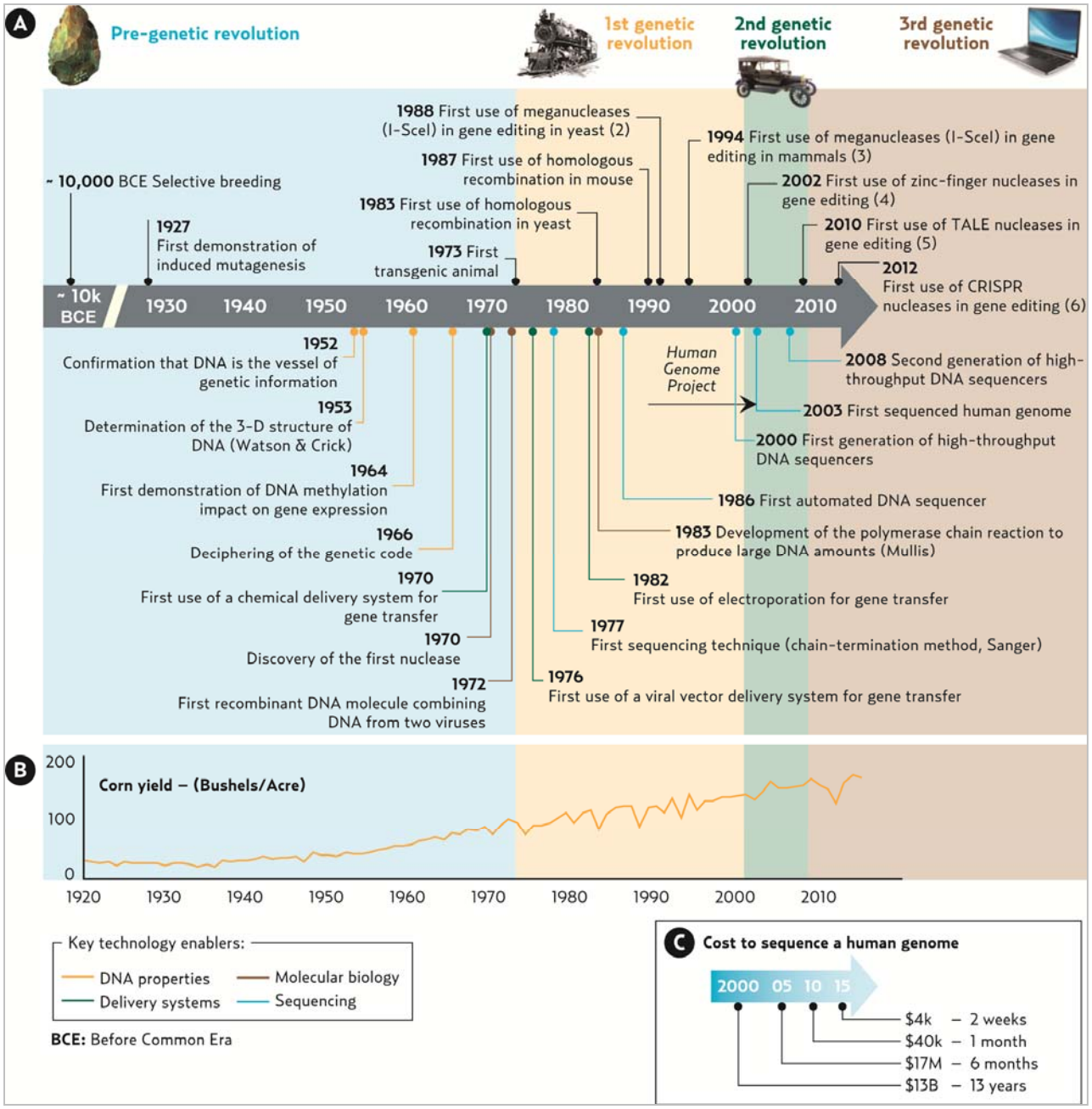


Figure 1. A brief history of genome engineering

A. Historical timelines of key events in genome engineering: progress has been enabled by a set of advances in the understanding of DNA properties, the ability to manipulate molecular DNA, to deliver it within cells and to sequence it.

B. Corn yield increase over the last century is shown as an example of the impact of advances in genetics.

C. Cost and time to sequence genomes has considerably decreased over the past 15 years.

[2] Rudin and Haber, Mol. Cell. Bio., 1988.
[3] Choulika et al., C R Acad Sci III, 1994.
[4] Bibikova et al., Gen. Soc. of Am., 2002.

1.2. …toward modern biotechnology

Modern biotechnology arose in the early 1970s with transgenesis: for the first time scientists were able to go beyond selective breeding possibilities by inserting external instructions to the genome of a species. This instruction is contained within a recombinant gene and confers a new characteristic; this was the start of **reverse genetics**. For example, the human growth hormone gene can be inserted in *E. coli* bacteria to produce this hormone in high quantities, which means that it does not need to be extracted from human tissues anymore, avoiding contamination risks such as prion contamination responsible for Creutzfeldt-Jakob disease.
The advent of reverse genetics also marked the entry into a period of continuous progress in understanding how the DNA content of genes shapes biological traits. This was enabled by a set of improvements in our capacity to **manipulate** DNA, to **deliver** it into organisms and to **sequence** it.

> **Classical genetics** starts from a biological attribute and seeks to determine the underlying genetic basis of this attribute.
> **Reverse genetics** is the study of gene function accomplished by analyzing the biological result of inserting the gene into an organism.

Transgenesis is the addition of a gene in the genome of a species, most often by the insertion of a DNA fragment at a random location in the genome.

This technology introduced the era of *genetically modified organisms*, from herbicide-tolerant plants to hamster cells producing human erythropoietin (EPO).[7] Most of the biological drugs today, such as monoclonal antibodies, are produced by cells modified by transgenesis.

It also led to the first clinical trials in gene therapy, aiming at treating diseases caused by a defective gene by inserting a functional one[8]. However, because the insertion is random, there is the risk that not only the transgene will malfunction but also that it will interfere with endogenous genes, potentially triggering harmful consequences.

> **DNA sequencing** is the process of determining the nucleotide sequence of a gene (or of any genomic region). Knowing the sequence to be modified is a prerequisite to gene editing.
> Advances made in this domain have been tremendous. Completed in 2003, the first sequencing of a human genome—the Human Genome Project—took 13 years and ~$3 billion. A human genome can be sequenced today in two weeks for several thousand dollars.

[5] Christian et al. Genetics., 2010.
[6] Jinek et al., Science, 2012.
[7] Erythropoietin (EPO) is a growth hormone that controls red blood cell production. EPO is used as a therapeutic agent to treat anemia resulting from chronic kidney disease or cancer treatment. EPO annual sales in 2014 amounted to $7.5 billion
[8] Cavazzano-Calvo et al. , Science, 2000.

Precise edition of a genome with controlled DNA modification at a targeted location was first performed in the 1980s by **gene editing** through **homologous recombination**.[9,10,11,12] It relies on the delivery into the cell of a DNA fragment carrying homologous sequences upstream and downstream to a target location. The information it contains is inserted into the genome at the target location (see Figure 2). Homologous recombination is at the origin of gene editing and the start of this revolution.

Homologous recombination was first described in yeast in 1983 by Jack Szostak[8] and colleagues. In the late 1980s, the work of Mario Capecchi,[9] Martin Evans,[11] and Oliver Smithies[10] (collectively awarded the Nobel Prize in 2007) led to the first knockout mouse using homologous recombination.

However, gene editing by homologous recombination has long been confined to academic labs due to its inefficiency: it occurs on average only once in every 1 million cells into which the DNA fragment is delivered when using mouse embryonic stem cells. In most other organisms and cell types, this frequency is even lower. Thus, whereas the knowledge in the field of genetics was continuously increasing, strongly driven by progress in sequencing, the translation of this knowledge into new therapies or other uses was limited by the ability to modify genomes efficiently and safely in the proper tissue.

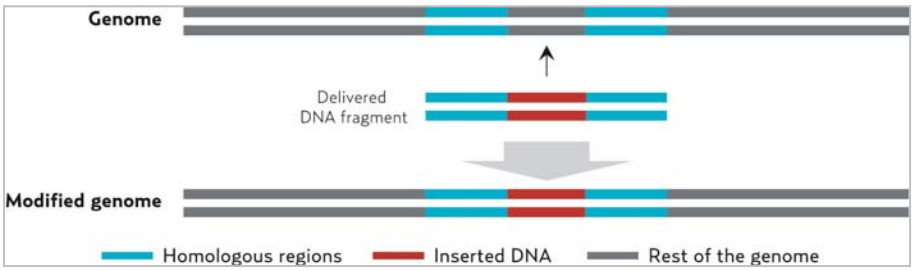


Figure 2. Homologous recombination

Thanks to the presence on the delivered fragment of DNA sequences homologous to the target location, the DNA information it contained is inserted in the genome at the target location.

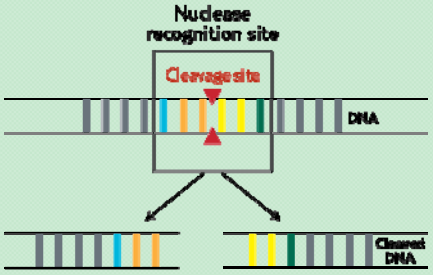
[9] Szostak et al., Cell, 1983.
[10] Thomas and Capecchi, Cell, 1987.
[11] Doetschman et al., Nature, 1987.
[12] Evans and Kaufman, Nature, 1981.

1.3. The gene editing revolution

The discovery that generating DNA breaks at the target location considerably enhances the efficiency of homologous recombination paved the way for **nuclease-based gene editing**. Nucleases are specialized proteins that recognize a specific DNA sequence and cleave it. Their use leads to a jump of homologous recombination efficiency of more than five orders of magnitude (driving it from 0.0001 percent to up to 20 percent in mammal cells[13]).

Even simpler, nucleases can also be used alone to edit genes (see Figure 3). In that case, DNA breaks generated by the nuclease are repaired through an error-prone process called non-homologous end-joining. This provokes DNA changes at the cleavage location, which most of the time result in the definitive inactivation of the surrounding gene (see Figure 4).

Nucleases are molecular scissors that cleave DNA at specific sites.



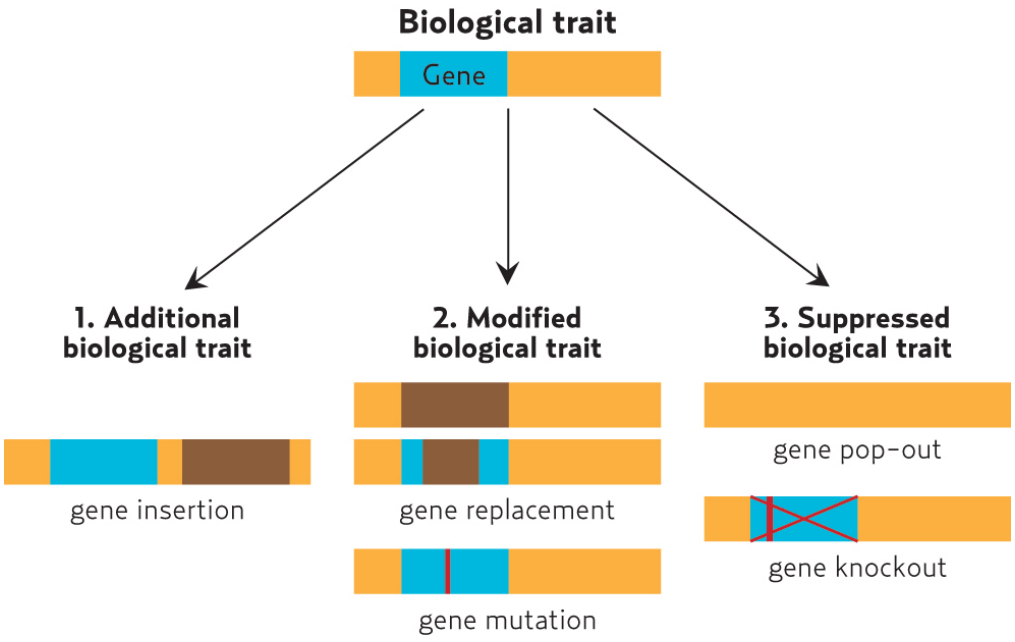


Figure 3. Three types of gene edition:

1. An additional biological trait can be obtained by inserting a new gene (gene insertion).

2. A biological trait can be modified either by i) the complete or partial replacement of the target gene (gene replacement) or ii) modifying its DNA sequence (gene mutation).

3. A biological trait can be suppressed either by i) removing the gene from the genome (gene popout) ii) or modifying its DNA sequence in a way that stops the gene from being functional (gene knockout).

[13] Voit et al., Nucleic. Acids. Res., 2014.

The first generation of nucleases, meganucleases and zinc-finger nucleases (ZFN), rose over 20 years ago. As previously explained, they were a huge improvement over gene editing by homologous recombination alone in terms of efficiency. However, they were costly and hard to engineer, requiring robots, expertise and fully dedicated labs, which prevented them from being widely adopted. A few years ago, TALE and CRISPR nucleases considerably improved the speed and cost of gene editing protocols. This is already seen as one of the most important technological breakthroughs since the Polymerase Chain Reaction (see Box) and ushers in a new genetic era that will have a huge impact on health, agriculture and beyond.

PCR (Polymerase Chain Reaction) is a molecular biology technology used to generate millions of copies of a particular DNA sequence. Developed in 1983, PCR has been a game changer for DNA manipulation in genetics labs and was awarded the Chemistry Nobel Prize in 1993.

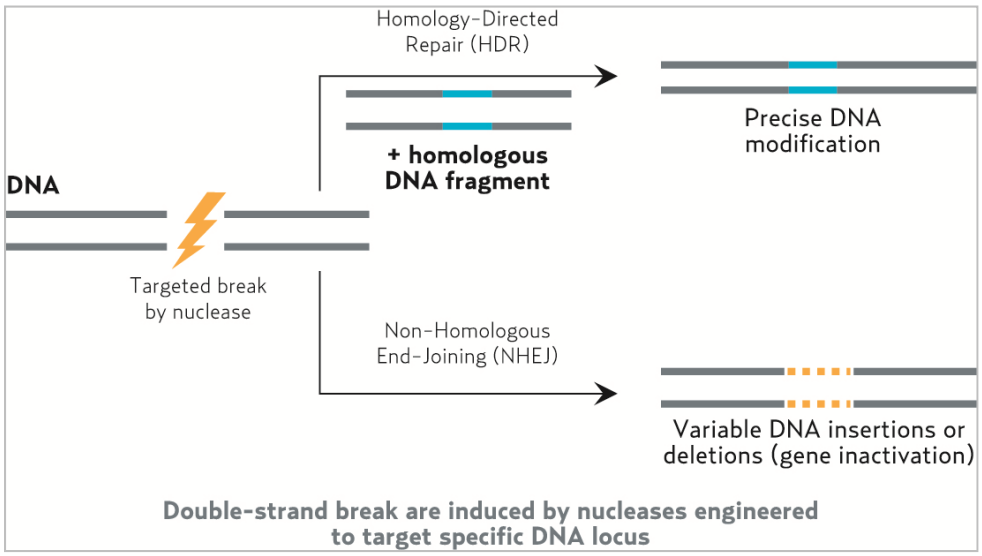


Figure 4. Nuclease-based gene editing can be used to modify DNA in two ways:

1. Homology-Directed Repair – a nuclease and a homologous DNA fragment are delivered to the cell. The DNA fragment is inserted into the genome at the location of the break, leading to a precise DNA modification.

2. Non-Homologous End-Joining – the nuclease is used alone. The DNA break is repaired through an error-prone mechanism, resulting most of the time in the inactivation of the surrounding gene.

1.4. Nuclease-based gene editing is already being used outside of academic research

Nuclease-based gene editing is already widely used in **research** as a cost-effective, fast, and easy way to conduct genetic experiments. For example, scientists inactivate or modify genes to understand their function and to generate cell and animal models that mimic a human disease. The latest technologies are so powerful that new research protocols aim no longer to modify one gene at a time but to alter the whole set of approximately 20,000 human genes in parallel and observe which ones are involved in a studied biological process.

More strikingly, nuclease-based gene editing is already being developed for human **therapeutics**, as a new approach to perform gene therapy: introducing or editing genes to cure a disease. Among the approximately 500 gene therapy clinical trials approved since 2010, only a few are based on nuclease-based gene editing, but many more are expected to be launched in the coming years. The most advanced nuclease-based gene therapy clinical trial is a phase II conducted by Sangamo Biosciences, using ZFN to modify immune system cells and prevent them from being infected with HIV. Sangamo is also collaborating with Biogen to develop ZFN-based therapies against hemoglobinopathies and with Shire in the domain of hemophilia, still this last collaboration very recently stopped, Sangamo pursuing on its own the development of this program.

Chimeric Antigen Receptor (CAR)-T cells are cells from the immune system engineered to express an artificial receptor at their surface. This receptor binds an antigen present specifically at the surface of cancer cells and consequently directs the immune system to "kill" the cancer cells, without toxicity for healthy cells. Nuclease-based gene editing is a new way to engineer CAR-T cells that is being undertaken by an increasing number of actors. Investments in this field have soared in 2014 and 2015 (see Figure 5).

Other very promising approaches are being developed in the field of immuno-oncology with "CAR-T cells", i.e., immune system T-cells engineered to fight cancer cells (see Box). While CAR-T cells are being developed since 2005, many Pharma and biotech players use transgenesis (e.g. Juno Therapeutics, Kite Pharma, Novartis, …), and most of these players are building partnerships with gene editing specialists since 2014 to use nuclease-based gene editing technologies to engineer CAR-T cells (see Figure 5). The main reason for this move is the ability to easily knock out selected genes with gene editing – hence silencing proteins that activate immune response. It paves the way for **allogenic CAR-T cell therapies** (using a unique T-cell bank vs engineering own patient cells for each therapy) with reduced risk of immune rejection and increased success rate of the therapy itself, Cellectis is the pioneer of this promising approach followed recently by most of the key "CAR-T cells" players.

Gene editing can obviously also be used to engineer the tumor targeting system of the CAR-T cells itself with a more specific and efficient approach vs. conventional transgenesis: some of the projects under development are relying on such "a pure gene editing approach". All of these new projects are today less advanced compared to those initiated earlier with transgenesis only, still the large amount of investments from most players in the field shows that nuclease-based gene editing is considered today as the most promising approach to accelerate the development of CAR-T cells as a safe and efficient cancer treatment.

Despite therapeutics getting much of the attention, it is likely that it is in the agribusiness industry that we will see the first marketed products. Following to a 20th century dominated by random mutagenesis induced by chemicals, the 21st century agricultural business will be marked by nuclease-based gene editing, which offers much more precise and controlled ways to modify critical traits in crops and animals. There is a great need to feed an ever-growing world population while limiting environmental

impacts. Key players in the plants field today include DoW Agrosciences and Calyxt, Cellectis' plant-dedicated subsidiary. DoW Agrosciences uses ZFN to develop herbicide-resistant corn and soybeans while Calyxt uses TALE nucleases to develop products such as high oleic soybeans or potatoes with better cold storage properties and lower acrylamide level. In livestock, Recombinetics has developed hornless dairy cattle using TALE nucleases.

The most important transformations are expected for the following years as more and more R&D programs are launched, building on the game-changing properties of the last generation of nucleases.

Deals in the field of CAR-T cells development

Company/ Institution	Date	Deal giving access to a nuclease technology?	Partner	Terms ($M)		Partner is a new entrant in CAR-T?	Most advanced on-going projects (not necessarily using nucleases)
				Upfront	Total		
Cellectis	Feb. 2014	Yes (TALE nucl.)	Servier	10	1 110	✓	
Cellectis	Jun. 2014	Yes (TALE nucl.)	Pfizer	80	2 800	✓	
Pregenen	Jun. 2014	Yes (MegaTAL[2])	bluebird bio	24	140		Preclinical
Memorial Sloan-Kettering	Sep. 2014	No	Atara Biother.	6	33	✓	
Transposagen Bio.	Nov. 2014	?[4]	Johnson & Johnson	Undiscl.	? (>584)	✓	
Oxford BioMedica	Nov. 2014	No	Novartis	14	90		Phase II - CTL019 for r/r CLL
Intellia Therapeutics	Jan. 2015	Yes (CRISPR)	Novartis	Undiscl.	Undiscl.		Phase II - CTL019 for r/r CLL
Kite Pharma	Jan. 2015	No	Amgen	60	1 110	✓	
MD Anderson	Jan. 2015	No	Ziopharm, Intrexon	115	175	✓	
Broad Institute	Jan. 2015	Yes (CRISPR)	AstraZeneca[1]	Undiscl.	Undiscl.	✓[1]	
Ohio State Univ.	Jan. 2015	No	Cellectis	Undiscl.	Undiscl.		Preclinical
Intrexon	Mar. 2015	No	Merck Serono	115	941	✓	
T-cell Factory B.V.	Mar. 2015	No	Kite Pharma	21	251		Phase I/II - KTE-C19 for r/a NHL
Crispr Therapeutics	Apr. 2015	Yes (CRISPR)	Celgene & GSK	64	64		Preclinical for Celgene[3]
Editas Medicine	May 2015	Yes (CRISPR)	Juno Therapeutics	25	? (>737)		Phase I/II - JCAR014 for B-cell malignancies
Five Prime Ther.	May 2015	No	bluebird bio	1.5	132		Preclinical
BioNTech	May 2015	No	Eli Lilly	60	360	✓	
Fate Therapeutics	May 2015	No	Juno Therapeutics	13	63		Phase I/II - JCAR014 for B-cell malignancies
bluebird bio	Jun. 2015	Yes (MegaTAL[2])	Kite Pharma	Undiscl.	Undiscl.		Phase I/II - KTE-C19 for r/a NHL
Juno Therapeutics	Jun. 2015	No	Celgene	1 100	?		Preclinical
Weill Cornell Univ.	Jun. 2015	No	Cellectis	Undiscl.	Undiscl.		

All CAR-T cells projects based on nuclease-based gene editing are in preclinical phase as of today

[1]AstraZeneca is not developing CAR-T cells for now but is highly expected to enter in the field soon given sustained interest in close areas and recent parnerships with Juno Therapeutics.

[2]MegaTAL are hybrid combinations between a meganuclease and a TALE nuclease.

[3]GSK/Adaptimmune are very active in another close field of T-cell therapy (engineered T-cell receptors) with NY-ESO TCR product in 2 phase I/II trials and 3 phase I trials.

[4]Main part of the deal is focused on Transposagen's piggyBAC transgenesis technology. However Transposagen is also developing nuclease-based gene editing technologies (including TALE and CRISPR nucleases) that may be also covered by the deal.

r/r ALL: relapsed refractory Acute Lymphoblastic leukemia, r/r CLL: relapsed refractory Chronic Lymphocytic leukemia, r/a NHL: refractory aggressive Non-Hodgkin's Lymphoma

IPO in the field of CAR-T cells development

Company	Date	Value
Kite Pharma	Jun. 2014	$134M
Bellicum	Dec. 2014	$161M
Juno Therapeutics	Dec. 2014	$265M
Cellectis	Mar. 2015	$228M
Conkwest	Jun. 2015	$173M

Tale nucl.
CRISPR nucl.
MegaTAL[2]
Other

Figure 5. Key deals and IPOs related to CAR-T cells development in 2014 and 2015 *(until end of June 2015)*

Since January 2014, almost all players in the field of CAR-T with projects under development – some of them already in phase I or II (Juno Therapeutics, Novartis, Kite Pharma) – have invested money to partner with a company with expertise in nuclease-based gene editing to combine these new technologies with CAR-T cells development approach.

2. Four nuclease technologies based on different mechanisms of action

Nuclease-based gene editing relies on the use of a nuclease to mediate a DNA break at a targeted location in order to generate a DNA modification. Each nuclease family is made of two core elements (see Figure 6): one that recognizes and binds the DNA target (the **DNA-binding element**) and one that mediates the cleavage (the **cleaving element**).

The prerequisite of any nuclease used for gene editing is to have a unique target site in the genome to avoid sheering the genome into pieces and inducing toxicity. To that end, nucleases used for gene editing have DNA-binding elements that recognize long target sequences: typically, an 18 nucleotide-long target sequence can be considered as unique as it is likely to occur only once for every 70 billion combinations (4^{18}), the equivalent of 20 times the size of a human genome (see Figure 7). A second factor driving the need for such long target sequences is that nuclease binding is not perfect: the DNA-binding element sometimes binds DNA sequences slightly similar to its designed target. This tolerance to a few mismatches with the target sequence increases the odds to cleave and modify other locations in the genome.

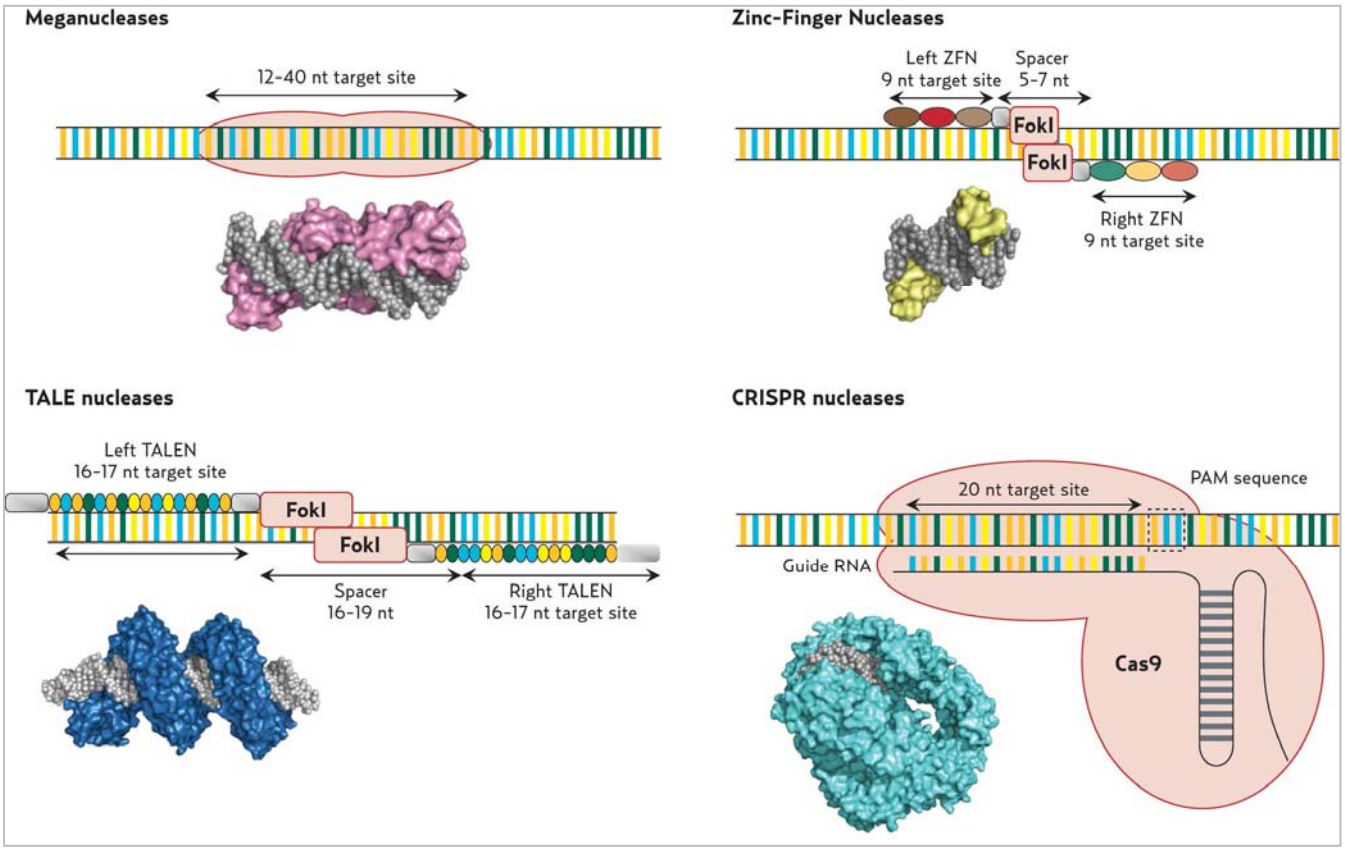


Figure 6. The four nucleases technologies: schemes and 3D molecular views

3D molecular views show the full molecule for meganuclease and CRISPR nuclease and only one DNA-binding element for ZFN and TALE nuclease.

FokI cleaving activity requires FokI dimerization. As a consequence, ZFN and TALE nucleases usually work as a pair.

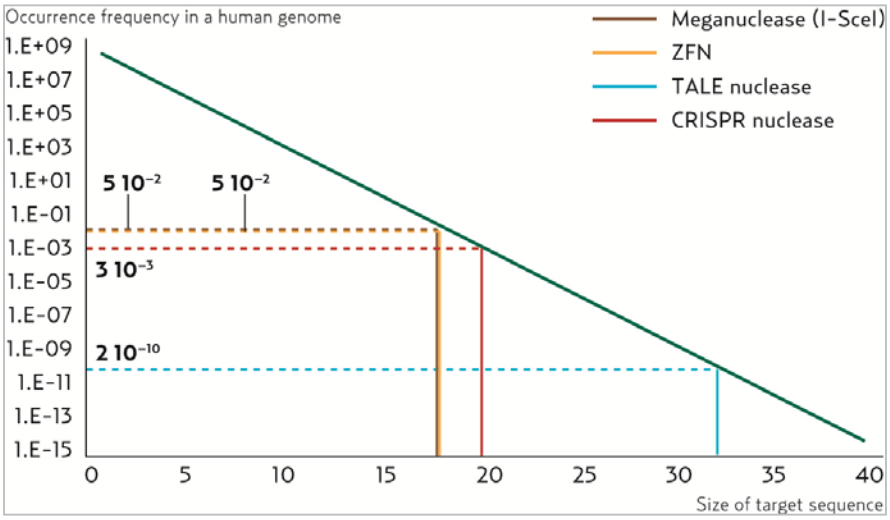


Figure 7. Relationship between the size of a target sequence and its occurrence frequency in a human genome

Longer target sequences are more likely to be unique. The typical target site occurrence frequency of each nuclease family is indicated. For example, the 18 nucleotide-long target sequence of the meganuclease I-SceI is likely to occur with a frequency of 0.05 in a human genome.

2.1. The first generation of nucleases was costly and hard to engineer

Meganucleases and Zinc-Finger Nucleases were the first tool developed for nuclease-based gene editing.

Meganucleases are derived from natural nucleases and characterized by long DNA recognition sites (up to 40 nucleotides) allowing them to be used as tools for gene editing, as such long sites are likely to occur less than once in a human genome on average. For example, I-SceI, the first gene editing nuclease and the best characterized meganuclease, recognizes an 18 nucleotide-long sequence.
The DNA-binding element and the cleaving element are entwined in meganucleases, which limit the possibility to design new meganucleases recognizing new targets (on average only one potential target every 1,000 nucleotides[14]). It requires a high degree of expertise and to perform laborious and expensive protein engineering and high-throughput screening.

Zinc-Finger Nucleases (ZFN) was the first solution developed to overcome meganuclease limitations in number of potential targets and ease of design. Developed in the 2000s, ZFN are artificial enzymes resulting from the fusion of two protein domains. The DNA-binding element comprises an array of 2-4 zinc fingers each recognizing a DNA motif. The cleaving element is derived from a nuclease called FokI. The combination of zinc-finger proteins within a ZFN can be tuned to recognize different target sites, allowing for a greater number of possible targets (around one every 140 nucleotides[15]). However, the efficiency of new ZFN nucleases is hard to predict as the DNA-binding capacity of each zinc finger is influenced by the others, requiring screening and testing many nucleases in parallel through long and expensive validation processes.

[14] Expert interview.
[15] Gupta et al., Nat Methods, 2012.

2.2. The second generation of nucleases is easy to use and much more flexible

TALE nucleases and CRISPR nucleases are two recently developed tools that have considerably broadened the ability to manipulate genomes sequences easily and effectively.

Transcription Activator-Like Effector (TALE) nucleases are artificial enzymes developed in 2009. Similar to ZFN, they are built from a fusion between a DNA-binding element consisting of an array of TALE subunits and a FokI cleavage element. The revolution brought by TALE nucleases relies on the modularity of the TALE subunits, each recognizing a specific DNA nucleotide, independently from the others. This one-to-one correspondence makes it very easy to design new TALE nucleases (about one week and a few hundred dollars) and to target any site in the genome.[16]

> Transcription Activator-Like Effector (TALE) proteins are derived from two plant pathogens (*Xanthomonas spp.* and *Ralstonia solanacearum*), which use them to target and activate specific host genes that will help the infection, hence their name. Recently, a DNA-binding protein cousin to TALE has been identified in a third plant pathogen (*Burkholderia rhizixinica*).

Finally, **Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) nucleases** are new tools, the development of which started in 2012. The strength of the CRISPR system is that the DNA-binding element is not a protein domain but a small RNA sequence called “guide RNA”. This guide RNA is very easy to design to target any new DNA sequence in the genome and also far simpler and cheaper to manipulate than proteins. This makes CRISPR a groundbreaking tool to perform gene editing in a few days and for less than two hundred dollars. This also requires very basic know-how in molecular biology and no specific expertise. The cleavage element is the nuclease Cas9, which can cleave any site in the genome as long as it contains a short sequence called PAM (occurring on average every 13 nucleotides in the human genome[17]).

> Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) nucleases are derived from bacteria immune system which uses CRISPR nucleases to degrade DNA of infecting viruses. The sequences used to produce the guide RNA against the different types of viruses are stored within the bacteria genome in a cluster of repeated sequences.

As previously mentioned, nuclease-based gene editing is increasingly used in a wide range of applications. The different nuclease technologies differ over a set of criteria, and those criteria have different relative importance depending on the application. This is expected to impact the rate of adoption of the different technologies as well as the future developments of the next generation of technologies.

[16] Reyon et al., Nat. Biotech., 2012.
[17] Le Cong et al., Science, 2013.

2.3. The criteria on which the nucleases technologies differ have different levels of importance in academic research and in therapeutics

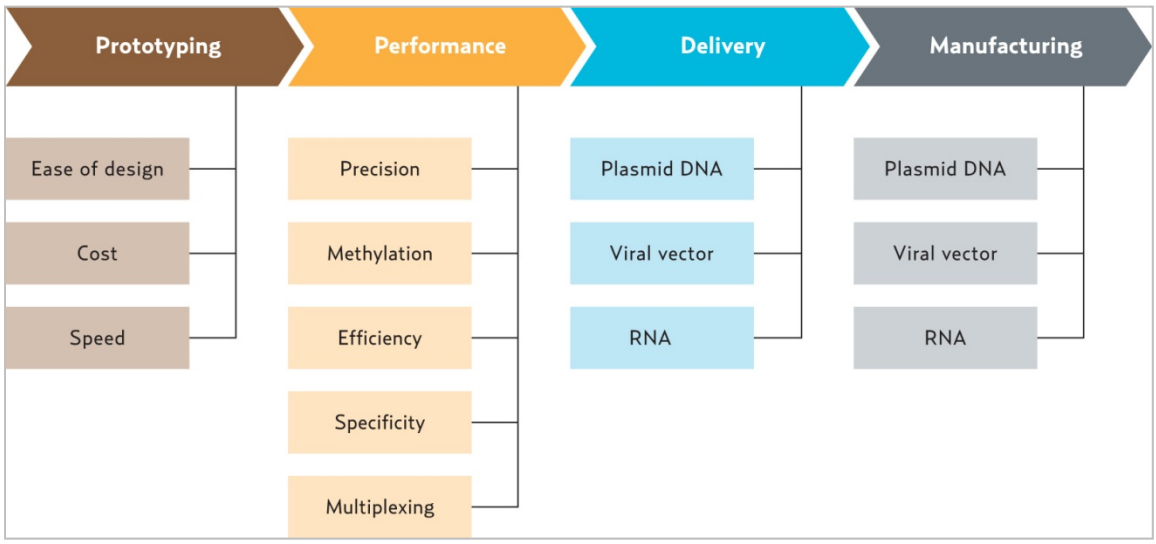


Figure 8. Four categories of criteria to consider…

… taking into account the critical steps of nuclease-based gene editing

The different nuclease-based gene editing technologies differ over a range of criteria classified into four categories: the prototyping of the system, the performance (including safety), the delivery, and the manufacturing (see Figure 8).

These criteria have different relative importance for the different application fields of gene editing. As an example, academic research and therapeutics, the two most important application fields today for gene editing, are characterized by highly contrasted needs.

Prototyping combines the level of skills, the time and the costs required to design and obtain a new effective nuclease. Prototyping is of prime importance for academic research as this step accounts for most of the time to perform an experiment and because research labs budget are constrained. On the contrary, new product development in therapeutics is very long and expensive, making the investment in nuclease prototyping marginal.

The **performance** of a nuclease technology depends on five parameters:

- Precision is the capacity to design a nuclease that can cleave as closely as possible to the desired location (see Figure 9). Precision is crucial in therapeutics where highly targeted modifications are looked for in order to use only the most efficient and safest ones. Most of nuclease-based approaches today rely on gene inactivation through the non-homologous end-joining mechanism. However, we are expecting to see more and more nuclease-based approaches relying on the simultaneous delivery of a homologous DNA fragment with the nuclease. This second approach will need an even greater level of precision (the more precise the technology is, the shorter the homologous DNA fragment can be: this facilitates delivery and enhances efficiency). The level of

precision required in academic research is usually less important; for example, gene inactivation can be performed by modifying genes at different locations or even within their vicinity.

- Capacity to cleave methylated targets: methylation is a DNA chemical modification that modulates the level of expression of a gene. Some nucleases are unable to bind a target if it is methylated, reducing the resulting precision of the nuclease family on methylated DNA sequences.
- Efficiency is the capacity for a nuclease to effectively cleave the targeted DNA site once it has been delivered into a cell. **Efficiency is crucial in therapeutics** as it is directly associated with the success of a gene or a cell therapy. In academic research, many experiments allow for the selection or the screening of modified cells post-experiment, reducing the need for very high cleavage efficiency.
- Specificity (off-target): nucleases targeting is not perfect and some nucleases can cleave off-target sequences that are slightly similar to their designed target. Specificity is **of prime importance in therapeutics to ensure safety of treatment**, as any DNA modification at undesired locations can, at best, alter the quality of target cells, and, at worst, have harmful consequences on patients health, such as activating genes involved in cancer. In research, the possibility to screen cells post-modification reduces the need for very low off-target activity. However, this additional step still represents extra work and time.
- Multiplexing is the capacity to generate several DNA modifications at a time within the same experiment, using several nucleases of the same family together. This is an interesting feature to study gene function in academic research as most biological processes are carried out by multiple genes. As of today, multiplexing is not a critical parameter in therapeutics: safely modifying one gene at a time is already a highly complex task. However, as techniques and protocols improve, we anticipate this criterion will become important.

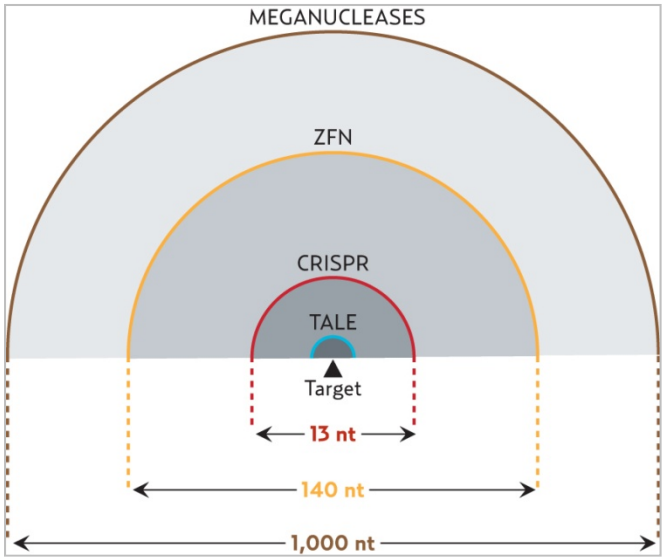


Figure 9. Theoretical maximum distance between a target site and the closest sequence recognized by a nuclease given the precision of each family

Note that when using some means to produce the guide RNA, it is necessary to have a specific nucleotide at one end, consequently reducing CRISPR precision.[18]

[18] Ranganathan et al., Nature communications, 2014.

The **delivery** of the nuclease system into the cell nucleus is mostly done via three vectors: plasmid DNA, DNA encapsulated in viral vectors, and RNA. These vectors carry the genetic information that will be used by the cell to produce the nucleases. In academic research, this criterion is of medium importance as researchers usually have the possibility to use different delivery systems for their experiments. In the gene therapy field, achieving gene delivery into the desired cell type remains one of the biggest hurdles to overcome and many efforts are dedicated to develop safe and efficient ways to achieve it. The delivery criterion is therefore capital for therapeutics.

As of today, the main vector type used for gene therapy is viruses. However, the delivery of a nuclease in a target cell is a hit and run procedure: once the target site has been cleaved and modified by DNA repair mechanisms, the presence of the nuclease in the cell becomes useless. Even worse, long-term presence of the nuclease is expected to generate high level of off-target cleavage resulting in high toxicity. Therefore, for therapeutic uses, in order to avoid side effects, delivery vectors resulting in transient presence of the nuclease in the cell will be preferred, such as RNA. The nuclease must disappear once the job is done. In addition, although not widely used today, delivery of the nuclease system as a protein[19] (or as a protein/RNA complex for CRISPR[20]) is expected to increase, as this vector allows both for the transient presence of the nuclease in the cell and for precise control over the amount of nucleases finally effective in the cell by eliminating the intermediate step of nuclease production by the cell.

Three main types of vectors exist to deliver nucleases:

Plasmid DNA is the basic form to handle DNA in molecular biology. Plasmids are small circular molecules carrying a desired DNA sequence.

Viral vectors harness viruses' evolution-shaped capacity to deliver their genes to cells in a pathological manner. Virus' DNA is modified to remove disease-causing genes and insert the desired DNA sequence.

RNA is a short DNA message molecule comprised of a chain of nucleotides. RNA genetic information is more quickly used by the cell to produce a nuclease and for a shorter period of time.

Manufacturing is the large-scale production of the vector chosen to deliver the nuclease system. High quantities (and high quality) are needed only for therapeutic approaches.

Because of their marked advantage in prototyping, precision and efficiency, TALE and CRISPR nucleases have emerged as the tools of choice for gene editing. Interestingly both are characterized by unique features and limitations.

[19] Luo et al., Mol Plant., 2015

[20] Schumann et al., PNAS, 2015

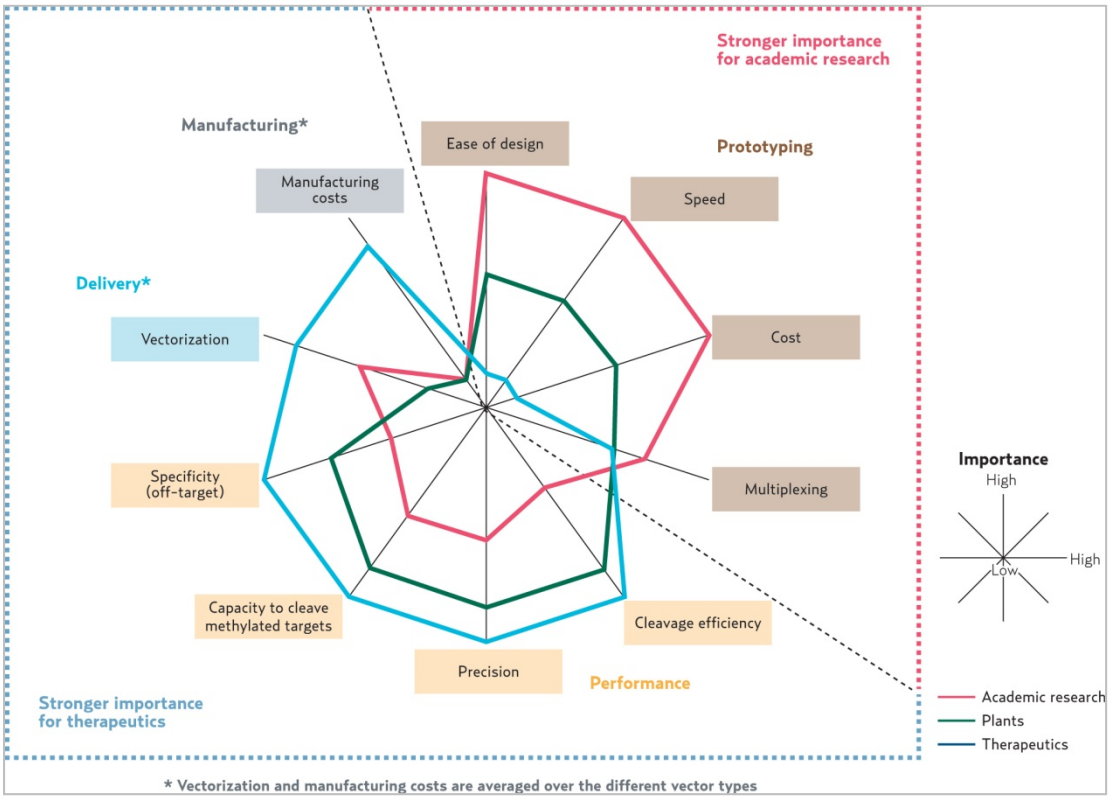


Figure 10. Relative importance of choice criteria in academic research, plants, and therapeutics

Plants application requirements are intermediate between research and therapeutics.

3. Focus on TALE and CRISPR nucleases: Two new technologies with distinct features and limitations

3.1. TALE nucleases: flexible and specific

TALE nucleases began to be developed in 2009.

TALE nucleases have been developed thanks to foundational work of teams led by Daniel Voytas,[21] Ulla Bonas,[22] and Adam Bogdanove[23] .Daniel Voytas has also been very active in the further development of the TALE nuclease technology, with more than 30 publications using this technology since 2010.

The high modularity of their DNA-binding element makes them highly flexible tools, able to target virtually any sequence in the genome. Among the four nuclease technologies, TALE nuclease **precision** is the highest (see Figure 9).

TALE nucleases are both **efficient** and **specific**. They exhibit the lowest level of off-target activity when cross-compared with ZFN and CRISPR nucleases.[24,25,26] For example, Valton et al.[27] managed to generate DNA modifications by RNA delivery at both the TRAC and dCK loci in primary T-cells at frequencies of 80% and 76% respectively without any detectable off-target activity.

The DNA binding capacity of the first developed TALE nucleases is sensitive to **methylation**. However, this drawback has since been overcome and it is now possible to design nucleases either sensitive or not to methylation,[28] providing enhanced flexibility to the user as it offers the possibility to discriminate between a methylated and an unmethylated allele.

TALE nucleases **prototyping** requires specific expertise and know-how in molecular biology and protein engineering: complex molecular operations are necessary because of the repetitive content of the DNA sequence encoding the TALE nuclease.

The repetitive content of the TALE nuclease sequence is also a limitation for the **delivery** with viral vectors (at least for lentivirus) because of the rate of recombination. In addition, the large size of the sequence encoding the TALE nuclease gene (2,800 nucleotides) and the necessity to deliver two components (see Figure 6) increase the complexity of delivery with the three types of vectors. It also translates into higher **manufacturing** costs for both RNA and viral vectors.

Interestingly, BurrH, a new modular DNA binding protein cousin to TALE, has been identified from the pathogen *Burkholderia rhizoxinica*[29]. BurrH nucleases could be an alternative to TALE nucleases as the sequence encoding BurrH does not show the same repetitive content found in the TALE domains.

Other recent developments have been conducted to overcome the need to deliver two components: MegaTALs and Compact TALE nucleases[30,31]. In these potential next generations, the FokI cleavage

[21] Christian et al., Genetics, 2010.
[22] Boch et al., Science, 2009.
[23] Moscou and Bogdanove, Science, 2009.
[24] Mock et al., Nucl Acid Res., 2015.
[25] Mussolino et al., Nucl Acid Res, 2014.
[26] Wang et al., Nature biotech, 2015.
[27] Valton et al., Mol Ther., 2015.
[28] Valton et al., J Biol Chem, 2012.
[29] Juillerat et al., Sci Rep, 2015.
[30] Beurdeley et al. Nature communications, 2013.

domain is replaced with a meganuclease or a meganuclease cleavage domain, allowing them to work as a monomer. As a result, MegaTALs and Compact TALE nucleases can be used as a single chain molecule, making them easier to handle and vectorize.

3.2. CRISPR nucleases: user-friendly and cost-effective

Although the development of CRISPR nucleases began less than three years ago, they have already been widely adopted by academic labs: the total number of publications mentioning CRISPR has already outnumbered the publications mentioning ZFN and TALE nucleases. In addition, in 2014, the number of CRISPR nuclease requests received by Addgene, a nonprofit organization that stores DNA constructs for academic uses, was ten times the number of TALE and ZFN nuclease requests. This fast and massive adoption is due to the **simplicity**, **speed**, and **low cost** to design a new CRISPR nuclease (requiring only a new guide RNA sequence) combined with a high **efficiency**. For example, Yasue et al.[32] reported DNA modification in one-cell mouse embryo at an outstanding 90% frequency using CRISPR nucleases to target the *fgf10* locus. In human primary T-cells, protocols are improving fast: while the first studies showed DNA modification at a rate of ~10% [33], Schumann et al.[34] recently managed to generate DNA modifications at the CXCR4 locus at a frequency of 55% by delivering RNA/protein complex and Hendel et al.[35] reached a DNA modification frequency of 94% at the CCR5 locus using simultaneously two chemically modified gRNA.

Multiplexing, which is easily achieved by using several guide RNA with the same Cas9 nuclease, is also seen as a powerful characteristic for some experiments. As an illustration of its potential, CRISPR technology can be used to perform genome-wide studies in cell lines: rather than inactivating one gene and observing which biological attribute is modified, scientists start from a biological attribute, such as resistance to a drug, and inactivate in parallel all of the 20,000 genes of the human genome to screen for the ones linked to the biological attribute. The main drawback to overcome is the constitutive expression of Cas9 with the consequent toxicity.

As of today, the main drawback of CRISPR nucleases is their **high level of off-target activity (low specificity)**, raising safety concerns especially regarding use in therapeutics. One factor explaining this high off-target level is that the CRISPR recognition site seems to allow for a high number of mismatches: recent studies suggest that only between 5 and 12 nucleotides in the guide RNA are really important for the binding of the nuclease system[36]. As mentioned previously, such short sites are far less likely to be unique in the genome (see Figure 7). In order to deal with this limitation, CRISPR-nickase or CRISPR-FokI, new generations of CRISPR nucleases, are being developed[37,38] . These new versions are modified to work as a pair, like ZFN or TALE nucleases, in order to increase the specificity by doubling the length of the recognition site. However, this comes at a price as paired nuclease systems are more complex to deliver. Alternatively, other approaches to reduce CRISPR off-target activity (i.e. enhance the specificity) rely on the optimization of the composition and the structure of the guide RNA[39].

[31] Boisset et al. Nucleic Acid Res, 2014.
[32] Yasue et al., Sci Rep., 2014.
[33] Pankaj et al., Cell Stem Cell, 2014.
[34] Schumann et al., PNAS, 2015.
[35] Hendel et al., Nature Biotechnol., 2015.
[36] O'Geen et al., Nucleic Acid Res, 2015.
[37] Mali et al., Nat. Biotechnology, 2013.
[38] Tsai et al., Nat. Biotechnology, 2014.
[39] Cho et al., Genome Res, 2014.

Delivery of single Cas9 nucleases is already difficult because of the large size of the DNA sequences encoding the Cas9 nuclease (4,200 nucleotides), in particular for viral vectors. Smaller versions of Cas9 are being developed[40] . On top of the size itself, the difference in nature of the Cas9 protein on the one hand and of the guide RNA on the other increases the complexity to deliver optimal concentrations of each component in the cell.

Finally, relatively important **manufacturing** costs are expected to be driven by the large size of the DNA sequence encoding Cas9 and the nature of the guide RNA.

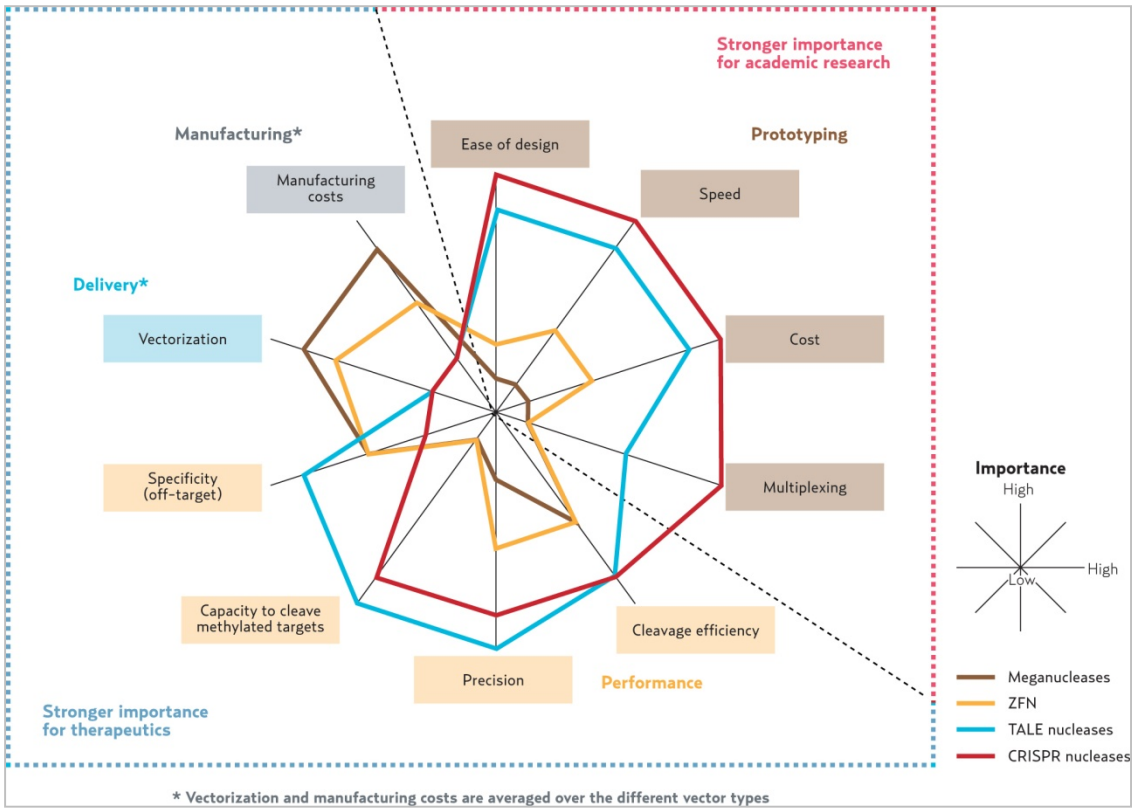


Figure 11. Comparison of the nuclease families' properties

Given the paramount importance of safety and quality control in therapeutics, the high off-target activity (low specificity) of CRISPR may limit its use in this application field, at least in its current version. Despite this potential drawback, the CRISPR technology has already attracted large investments and partnerships from big pharmaceutical companies to develop new therapeutic approaches based on CRISPR. An explanation for this might be found looking at the CRISPR intellectual property landscape which is not yet settled. This attracts many initiatives to grab part of this IP landscape, fueled by the absence of a high knowledge barrier to entry.

[40] Ran et al., Nature, 2015.

3.3. The intellectual property landscape of the new generation of nucleases

The nuclease-based gene editing revolution is also visible looking at the increasing intensity of the intellectual property activity (see Figure 12). Over the last five years, there has been a strong increase in the number of patent applications for nuclease technologies with an average annual growth rate exceeding 40 percent, driven by applications related to TALE and CRISPR technologies.

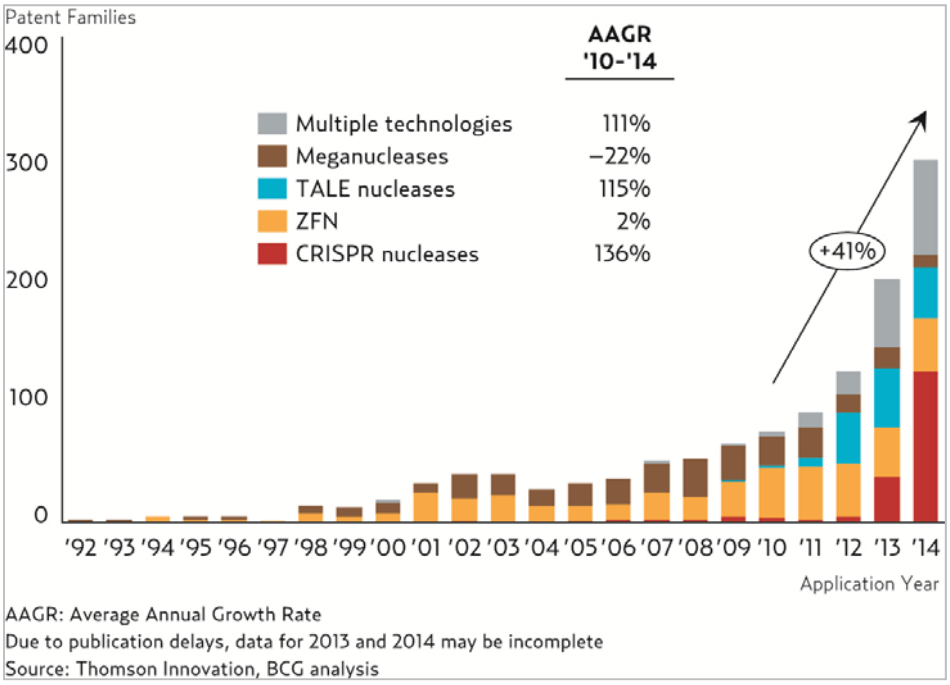


Figure 12. Number of patent applications filed per year for each nuclease technology

"Multiple technologies" bucket accounts for patents relating to at least two nuclease technologies.

Interestingly, TALE and CRISPR technologies show contrasting intellectual property landscapes, bearing some resemblance to the Apple vs. Android operating systems in mobile phones: the TALE nuclease technology IP landscape is well defined and consolidated around Cellectis (similarly to Apple) whereas the CRISPR IP landscape is more scattered and blurry. This creates an apparent "freedom to operate" which, combined with the high accessibility of the CRISPR technology, may recall the Android situation. On top of that, a legal dispute among the key stakeholders over the CRISPR intellectual property may prevent any clarification in the near future.

The TALE nucleases IP landscape is characterized by the strong and recent patenting activities of Cellectis, accounting for 13 percent of the total number of patent applications (see Figure 13). The Universities of Iowa and Minnesota are also key players. In addition, Cellectis has consolidated its position by licensing two foundational patents for the TALE nucleases technology. One patent filed by the Universities of Iowa and Minnesota[41] licensed to Cellectis in January 2011, and another one from

[41] WO2011072246.

Martin Luther University[42] in Halle, Germany, first licensed to the Two Blades Foundation and Life Technologies in 2011. Thanks to licensing deals in 2014 with these two actors, Cellectis obtained rights on this second patent and secured its position as the main actor for the commercial use of TALE nucleases in plants and therapeutics.

Interestingly, Sangamo Biosciences, the leader in ZFN, was involved in the early development of the TALE technology. However, Sangamo Biosciences seemingly did not pursue this approach after 2013.

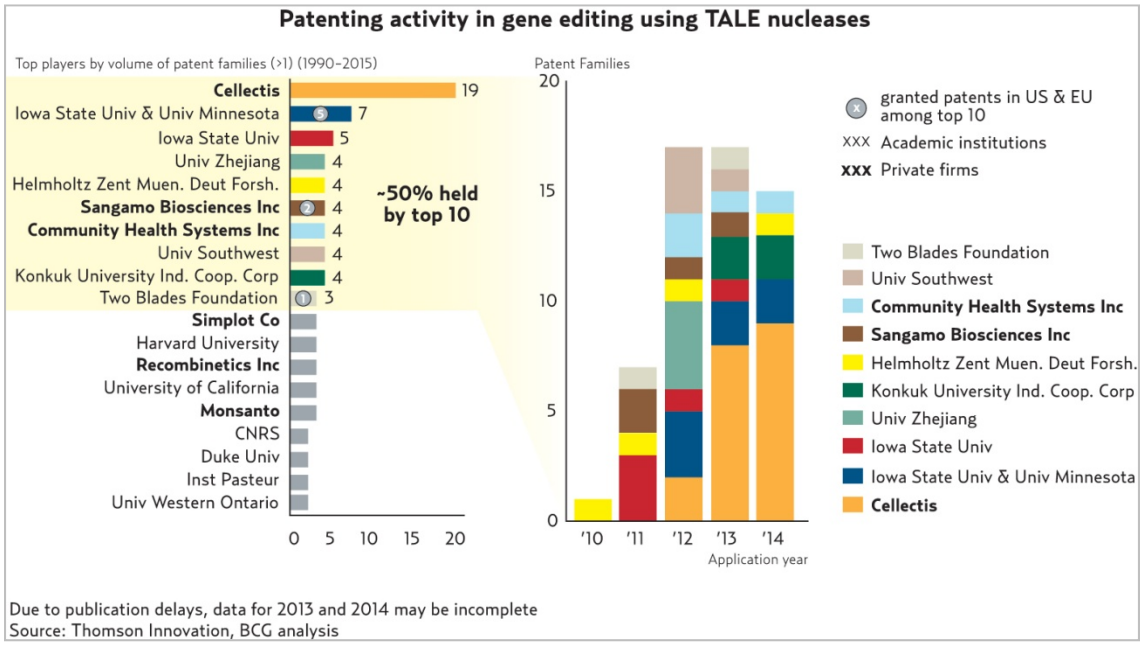


Figure 13. Top players patenting activity in the TALE nuclease field

Patents related to more than one nuclease technology are not included.

The CRISPR IP landscape is very recent, with the first applications related to gene editing filed in 2013 (prior filings are related to an alternative use of CRISPR, see Box). The patenting activity is largely dominated by academics, in particular Harvard University and MIT, under the auspices of the Broad Institute, with many co-filings between these institutions.

> In its natural form, CRISPR is a part of the adaptive immune system of bacteria. Each time a bacteria fights a virus infection, it stores in its genome a guide RNA to fight any further infection from the same virus more efficiently. Danisco has patented and exploited this process since 2007 to "vaccinate" bacteria used to produce yogurt and cheese.

[42] WO2010079430.

One of the most central patents for gene editing with CRISPR nucleases has been **filed** by UC Berkeley in March 2013. However, the first patent **granted** for gene editing with CRISPR is owned by the Broad Institute, filed six months later but under a fast-track process. In April 2015, UC Berkeley filed a request for "patent interference", asking the U.S. Patent Office to reconsider its decision. This has been the start of a legal dispute that may take years to be settled.

Despite this scramble, three companies founded in the last two years are competing today, using these patents to develop therapeutic approaches: CRISPR Therapeutics, Editas Medicine, and Intellia Therapeutics (see box). In total, they raised more than $330 million in venture capital since 2013.

Emmanuelle Charpentier and Jennifer Doudna are two inventors of the pending patent from UC Berkeley involved in a legal fight against the Broad Institute. The Broad Institute owns the first granted patent. Feng Zhang is one of the inventors of this granted patent.

CRISPR Therapeutics (Basel, Switzerland) was cofounded in Nov. 2013 by E. Charpentier. It raised $89 million in series A & B VC financing.

Editas Medicine (Boston, US) was cofounded in Nov. 2013 by F. Zhang and J. Doudna. It raised $163 million in series A & B VC investment. J. Doudna left to back the creation of Intellia Therapeutics.

Intellia Therapeutics (Cambridge, US) was cofounded in November 2014 by J. Doudna. It raised $85 million in series A and B VC investment.

In addition to these core patents, other patents may prove to be essential in the use and commercialization of CRISPR technology and are already in demand. As an example, Dupont Pioneer, a leading seed company, entered an exclusive license agreement with the University of Vilnius. Similarly, Cellectis has exclusively licensed the main patent family from the University of Minnesota, entitled "Engineering Plant Genomes Using CRISPR/Cas Systems."

Beyond the patents covering the use of the different nuclease technologies, upstream patents required to perform gene editing operations should also be considered as they might disrupt nuclease technology owners' freedom to operate. In particular, homologous recombination is expected to be increasingly used with nucleases to introduce more complex DNA modifications than can be obtained with the NHEJ pathway. Interestingly, in January 2015, Cellectis announced the issuance in the U.S. of a patent covering the use of chimeric nucleases for gene editing by homologous recombination.[43] This would encompass ZFN, TALE nucleases and megaTAL, some engineered meganucleases and potentially CRISPR nucleases.

[43] US 8921332.

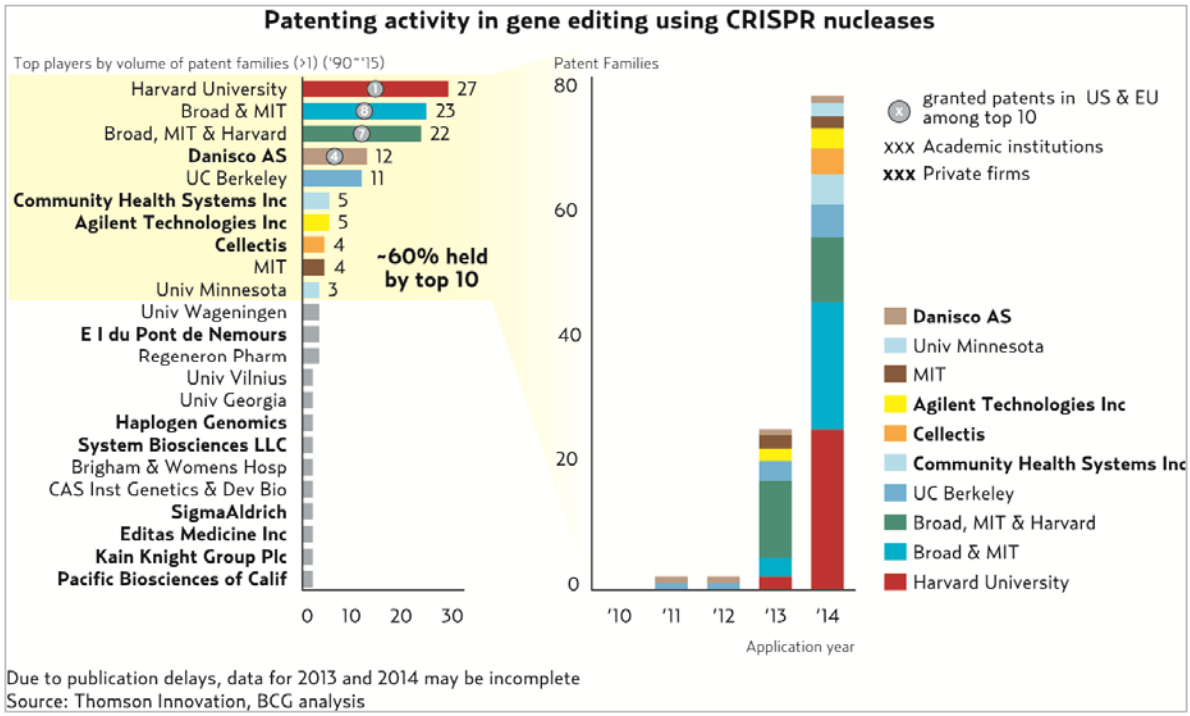


Figure 14. Top players patenting activity in the CRISPR nuclease field

Patents related to more than one nuclease technology are not included.

4. Many developments to follow in the coming years and beyond

We are expecting many developments in the upcoming years. The race has already begun in the immuno-oncology field, with impressive amounts of capital invested. Sangamo's approach to fighting HIV infection by modifying T-cells is another example of the gene editing potential. And, if we let our imagination run free, we can imagine a not-so-distant future in which nuclease-based gene editing treatment would be prescribed by your GP as commonly as antibiotics today.

More disruptive changes can be anticipated, in particular in combination with other improvements in sequencing, DNA, RNA and protein synthesis, or stem cells. Scientists have already used CRISPR nucleases to engineer human embryos[44] (these embryos were non-viable prior to experiment, they were obtained from a fertility clinic and intended to be discarded), sparking ethical debate all over the world. Opponents to the use of gene editing in human embryos worry about the unpredictable effects it could have on future generations, as well as the risk of exploitation for non-therapeutic modifications (such as choosing the eye color or IQ of a child)[45]. Others believe that gene editing in wild animals should be regulated, to avoid irreversible damage to ecosystems[46]. On the other side, advocates of gene editing research on human embryos underline the potential as a therapeutic tool, as 6 percent of all births have serious birth defects with a partial or exclusive genetic origin[47]. They consider that the risk of non-therapeutic uses should not justify restrictions on therapeutic uses (comparing it with the internet and the risk of cyberterrorism) and explain that further experiments will improve the safety of the techniques. In response, the U.S. National Academy of Sciences is launching an international summit in fall 2015 to explore the ethical issues with human gene-editing research, reminiscent of the 1975 Asilomar Conference, which discussed the first genetic modifications of bacteria. In this area, sci-fi literature and movies have already suggested many possibilities about what could be life with gene editing. Nevertheless, 40 years after Asilomar, genetically modified bacteria have proven to be fantastic tools to improve human health. Let's bet that these gene editing technologies will bring also such positive impacts.

With an increasing need to produce "more and better" while preserving the environment, agriculture is another field of expansion for gene editing where it is used to improve attributes such as yield, pathogen resistance, herbicide tolerance, nutritional properties, or allergen content. Beyond dealing with these existing needs, we can imagine a future in which gene editing will be directed by Chefs to create new foods with new tastes or properties.

Bioproduction, i.e., the use of modified organisms as living factories, will continue to expand thanks to gene editing. The number of therapeutic proteins that are produced by cell lines will increase and the manufacturing costs will decrease. Biofuel will be produced on a large scale from organisms that can use sources not competing with food supply. And improvements for every type of material with a biological origin can be considered: wool, leather, wood, rubber, organic oil. We can imagine that in the future each home will have a "cell factory" that will be used similarly to a 3D printer, with specific "edits programs" to produce desired biological products, such as fuel for the car. Even light can be produced by organisms: in June 2013, a California startup company raised $500,000 in crowdfunding to design trees that can glow in the dark thanks to the addition of a gene from luminescent marine bacteria.

[44] Liang et al., Protein & Cell, 2015.
[45] Lanphier et al., Nature, 2015.
[46] Lunshof, Nature, 2015.
[47] Savulescu et al., Protein Cell, 2015.

Finally, as a sign that gene editing will have become a commodity, we can anticipate its use for non-productive purposes. Today, transgenesis is already used by some for entertainment and bioart, such as the creation of glowing fish, blue roses, or modified bacteria to create black and white photographs. In this field, we can let artists dream about a future in which gene editing will serve their creativity and their imagination!

***

About the Authors

This report reflects BCG's deep expertise in the biopharmaceutical industry, and in particular, on the topic of research & scientific development. The research for this working paper was supported by BCG's Health Care practice and case work with clients with active interests in the gene editing field. Please contact the authors for further discussion on the topic.

Elsy Boglioli is a Partner and Managing Director in the Paris office of The Boston Consulting Group. She is a member of BCG's Health Care and Strategy practices. You may contact her by e-mail at boglioli.elsy@bcg.com.

Magali Richard, PhD is a Project Leader in BCG's global Health Care practice and based in Paris. You may contact her at richard.magali@bcg.com.