

EDITING THE HUMAN GENOME

role in functional genomics and translational medicine

It is now readily accepted that we are in a post genomic era. With the steady flow of genomic information available to researchers worldwide, the focus turns to ways to analyse this information effectively and then utilise it in a practical manner.

Despite having cracked one layer of the genetic code (DNA), the complexity of the biology that exists in a single mammalian cell is still overwhelming. We are still dealing with the intricacies of post translational modifications, epigenetic changes associated with chromatin structure and transcriptional and post translational mechanisms which all contribute to this enormous complexity. Given that we do at least have an understanding of the DNA code underlying many of these processes, a necessary part of the experimental work that needs to be done calls for the ability to make changes to the genetic code and investigate what effects these changes might have. We must manipulate the DNA of living cells in its natural state to take full advantage of what we have learned from the sequencing of the genome.

The field of mammalian gene editing got its start in mouse ES cells and in 2007 Capecchi, Evans, and Smithies were recognised by the Nobel committee “for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells”. Unfortunately, altering the sequence of endogenous genes within differentiated human cell types has proven to be orders of magnitude less efficient than in mouse ES

cells. The ability to routinely and accurately edit the genome of somatic mammalian cells has required a healthy dose of innovative technology development. Persistence and hard work have paid off and now researchers find themselves armed with a number of tools which allow editing of the genome of living cells at acceptable frequencies.

Modern gene-editing technologies

Modern methods of gene editing currently fall into two broad categories. The first relies exclusively on homologous recombination, a natural DNA-repair mechanism, to perform endogenous DNA alterations and is best exemplified by the use of recombinant AAV (rAAV) as a gene editing tool. The second category functions through the stimulation of locus specific DNA repair events as a consequence of introducing double strand DNA breaks and is best exemplified by zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and meganucleases. Each approach has relative advantages and disadvantages depending on the desired end result. The focus of this article is not to mandate any particular approach as superior to any other, but to point out the variables which should be taken into consideration when designing gene editing projects.

**By Dr Chris Torrance
and Eric Rhodes**

Genomics

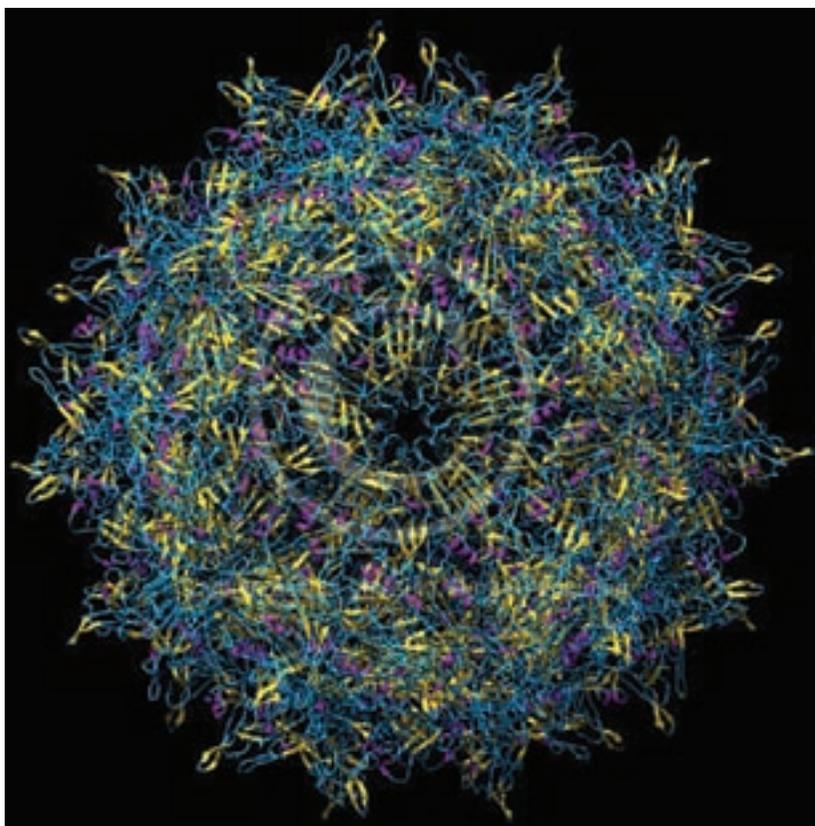


Figure 1
Example of an adeno-associated virus

Homologous recombination-based methods

As described above, this method is simplistic in its approach of providing a template DNA to the cell and allows the homologous recombination (HR) machinery of the cell to incorporate it at the site of homology. The length of homology used and whether the donor DNA is provided in single or double-stranded form can have a large impact on the efficiency of HR.

Traditional dsDNA approaches: Similar to the approach taken by the Nobel laureates in their work in mouse ES cells, this approach requires the introduction of a large piece (to achieve even passable recombination frequencies) of homologous DNA (generally in plasmid form) into a cell and subsequent screening for the desired HR event. The frequency for such events in typical mammalian somatic cells is on the order of 1×10^{-6} , rendering this a non-viable approach in most situations.

rAAV: Recombinant adeno-associated viruses are the dominant form of HR dependent gene editing (Figure 1). rAAV is a non-pathogenic single stranded DNA-virus that has a unique and powerful capability to induce HR at efficiency rates $\sim 1,000$ -fold greater than those seen using simple double-stranded vectors. It is not entirely known why they are so

efficient, but it appears that a distinct form of DNA-repair operates to faithfully recombine ss-DNA species into target genomic loci, which is independent of many of the factors typically seen to be important for dsDNA-mediated HR, eg Rad51 and Rad54b. Another advantage of using rAAV is its wide tropism. With more than 10 natural serotype variants and a number of chimeric capsids in use, rAAV is able to efficiently transduce greater than 90% of all mammalian cell types without the need for harsh electroporation or transfection conditions which are required by most other approaches. Finally, because rAAV is wholly dependent on homologous recombination, it is a very precise technique for introducing genetic modifications and does not suffer the off-target issues seen with other approaches. One current disadvantage of rAAV, however, is that targeting of alleles is generally done sequentially and so targeting of multiple alleles can take additional time compared to nuclease methods.

Nuclease-based methods

Although there are several different nuclease-based approaches, the mechanism by which gene editing occurs is shared among them. Upon introduction of a dsDNA break, the cell attempts to repair the break and has two main pathways that it can invoke, non-homologous end joining (NHEJ) and homologous recombination (HR). NHEJ is an error prone repair mechanism in which the cut DNA ends are quickly joined back together, often with minor deletions or additions at the break site due to the action of repair enzymes. This can be exploited to introduce disruptions in the coding sequence of a gene, but the end result cannot be predicted *a priori* and one has to screen clones to identify an NHEJ event which gives the desired result. The HR repair mechanism is similar to that of the HR directed process described above, however with nuclease-based methods the efficiency of HR is often highly dependent upon having the dsDNA break site close (<100bp) from the intended site of modification. In each cell line there is a balance between NHEJ and HR which plays out in determining how a dsDNA break will be repaired. There are currently no proven reliable methods to influence which repair path a cell will take and so the investigator is left to determine the nature of any dsDNA break repair which occurs.

Meganucleases: These are based on a category of naturally occurring enzymes which tend to bind to rather lengthy binding sequence (12-40bp) and cause double stranded breaks at the site of binding. Once considered the gold standard for specificity, recent reports have revealed that these enzymes do suffer significantly from off-target binding and

cleavage. Several groups have undertaken efforts to engineer meganucleases to bind to novel target sites, but the flexibility of this approach has been limited, primarily due to the complex interactions between the DNA recognition domain and nuclease cleavage domain. The enzyme is generally introduced into a cell encoded on a plasmid via electroporation or transfection. Meganucleases have seen limited use in engineering of mammalian cells.

Zinc-finger nucleases: These are hybrid vectors that combine an adaptable, sequence specific zinc-finger DNA-recognition domain fused to a dimerisation-dependent nuclease, usually FokI. When two zinc-finger nucleases (ZFNs) co-locate at a bipartite recognition sequence they create a dsDNA break. The design of ZFNs with high specificity has been the most challenging aspect of this approach. Although several public sources for ZFN module design exist, most reports indicate the designs obtained through these methods are inferior to commercially available sources which utilise proprietary design algorithms. The principal drawback to ZFNs

is the unpredictability of off-target cuts occurring in the genome and this issue is only amplified when using inferior design or selection strategies.

TALE-nucleases: This is a relatively new nuclease player in the field. Derived from a plant pathogen, TALENs are almost completely modular and deterministic in their assembly, allowing a simple design approach. Public sources for TALEN design algorithms also exist and it would appear that the only advantage of using commercially available sources is avoiding the labour-intensive part of assembling a full TALEN. The specificity of the TALEN modules is still being investigated and early reports indicate specificity roughly similar to ZFNs, but whether that specificity will hold up across a wide array of sequences has yet to be demonstrated. TALENs suffer the same issues as ZFNs with regard to predictability of potential off-target cutting.

Applications of gene editing

Given the variety of methods and associated advantages and disadvantages, there is no one



What — You *Don't* Dip and Read™?!

Do less and get more from your ELISAs

- Quantitate 96 protein samples in 15 minutes, not hours
- Take advantage of a 4-log dynamic range and dilute less
- Run automated assays without wash steps or overnight incubations
- Analyze crude samples with less prep
- Get information-rich direct binding data, not end-point results
- Develop new assays in a few hours, not weeks



The Octet® Platform

**Octet Systems with Dip and Read assays.
No hassle, label-free ELISAs.**

fortéBIO®
A Division of **Pall Life Sciences**

PALL® Life Sciences

fortebio.com • 1-888-OCTET-75

Fast. Accurate. EASY.

Genomics



Figure 2
Genetically engineered human cancer cells being grown in liquid tissue culture

method which serves all purposes best. The key differential lies in whether use of an HR versus NHEJ approach provides a distinct advantage. Another factor is ease of use and how the work is to be performed. When choosing among the various nuclease approaches, it is more a matter of cost and reliability that come into play. Let's entertain some typical examples of gene editing projects and discuss the pros/cons of the various approaches.

Example 1: Simple gene knock-out

Faced with determining the role of a particular kinase in a biological pathway, an investigator determines that due to the enzymatic activity of the gene, a simple siRNA knock-down is not providing her a clear picture. The investigator decides that a functional knock-out of the protein will serve her best. Which approach should she take?

Generally the first question that comes to mind is what is fastest (and cheapest – we will address cost issues later)? Due to the sequential nature of using rAAV, the simple answer to that question will generally be a nuclease-based approach for simple knock-out. But there are several important factors she should evaluate prior to settling on an approach.

1. First of all, what cell line does she want to work in? Is the cell line easily electroporatable or transfectable? If not, nucleases may be ruled out and rAAV becomes the choice.
2. How many copies (alleles) of her gene are there in her cell line? Biology 101 tells us that each human cell has two copies of each gene, but many of the most common immortalised lines have chromosomal duplications or deletions. If there are more than two alleles, then a nuclease approach may serve best.
3. Is the protein part of a gene family with consid-

erable homology among the family members? If so, the nuclease approach is at higher risk for off-target activity and screening for these events might require considerable effort. rAAV might be a better approach.

4. Does the kinase protein perhaps serve a scaffold function separate from its kinase function? Use of a commercially available nuclease would tend to target the first third of the coding region in hopes of introducing a shift in the coding frame that may or may not lead to expression of a partial protein. If scaffolding is not a consideration, then nucleases are well suited, however, if scaffolding is a consideration, then rAAV would be ideal for creating a kinase-dead mutant which still retained scaffolding function.

Example 2: Simple knock-in

Let's look at another investigator who wants to introduce an oncogene activating mutation into a gene. The same first question always comes to mind. What is fastest? Because this is a single allele targeting event, there is no immediate difference between rAAV and nucleases, so this is not quite so easy to answer in the case of a knock-in. But let's look deeper:

1. What cell line does he want to work in? In the case of the use of a nuclease, not only will delivery of the plasmid encoding the nuclease be required, but delivery of a second donor plasmid will also be necessary. If the cell is not easy to electroporate or transfect then rAAV would be a safer bet.
2. What is the nature of the knock-in? Is it a single base change or are there multiple changes over a broader area that need to be introduced? The efficiency of introducing changes by HR using nucleases is highly dependent on how close the desired change is to the DNA break point. If there are spatial changes involved (spacing of changes more than 20bp apart) then rAAV is probably the most reliable method.
3. Are my cells going to favour NHEJ over HR? While difficult to determine *a priori*, use of nucleases may be confounded by the prevalence of NHEJ and lack of HR activity. Using rAAV, which only works through HR, this issue can be avoided.
4. Is the area he wants to modify homologous with any other genes? Due to the need for the nuclease cut site to be proximal to the knock-in site, nuclease design strategy may favour design towards regions which lie further away from the desired target and lead to much lower efficiencies of HR. Because the region of homology used in rAAV targeting is so large (generally about 3Kb) it is easily able to distinguish between minor regions of homology which may exist.

Efficiency, time and cost

The one issue related to timelines which all of these gene editing approaches share is that none are efficient enough to yield a high percentage of modified cells in one treatment to allow the use of pools of cells. In each case, whether it be a knock-in or knock-out, single cell clones must be isolated from the bulk population, confirmed for the positive gene editing event, and then grown up in sufficient numbers to allow experimentation to occur. Different cell types grow with different doubling times and so each approach is limited by this one essential requirement. The time for creation of a nuclease reagent (at least in the case of ZFNs and TALENs) or the creation of an rAAV targeting vector is comparable and ranges from 6-10 weeks on average from commercially available sources. So except for situations which require multi-allele targeting, ZFN, TALEN and rAAV approaches would be roughly equivalent. In the case, however, of multi-allele approaches, nucleases do provide an advantage over rAAV in terms of time.

Cost structures for the various approaches vary widely and are dependent to a high degree on the level of validation or service component associated with the work. ZFNs, TALENs and meganucleases are commercially available as reagents (Sigma-Aldrich, Life Technologies and Collectis) and it is the investigator's job to undertake the editing and screening work herself. As of this writing Sigma was offering a limited custom gene editing service for ZFN work. rAAV comes with a unique arrangement whereby academic and non-profit organisations are provided with the necessary reagents and protocols at no cost to carry out editing projects, with any resultant lines flowing back to the commercial organisation (Horizon Discovery) for distribution. For-profit organisations can avail themselves of a dedicated and rapidly expanding custom rAAV gene editing service from Horizon.

Where do we go from here?

The work initiated by the 2007 Nobel laureates opened up a new world of functional genomics in mice and researchers have gained invaluable insights through use of those model systems. But now we have the tools to engineer human cell-lines. We must take full advantage of this capability and use it to functionalise human genes to better understand their role in the context of human biology and disease. Research in certain disease areas, such as oncology, which has such a large genetic component, will benefit enormously from the use of genetically defined human disease models (Figure 2).

Is the apparent slow adoption of these technologies perhaps due to a lack of awareness or fear that they are still too unwieldy? This review has attempted to provide some level of awareness and hopefully allay some of this fear while highlighting the capabilities that already exist. Modifying endogenous human genes is within the grasp of any lab now, and there are highly active support networks available (eg the ZF consortium, rAAV.com, and Horizon's Centres of Excellence programme).

We must not fail to grasp what is now within reach. The natural follow-on to the success of the Human Genome Project and the ENCODE Project is to implement that knowledge in a Translational Genome Project. Large scale endogenous genome editing efforts should be initiated to understand the complexity of the genome in multiple tissues and cell-types. We should develop a broad range of human disease models that faithfully recapitulate predisposing or pathogenic genetic variations (SNPs and mutations, respectively) which will sit in a continuum along with established genetically-engineered mouse models. This effort can be melded with on-going work in developing iPS and patient derived materials to accelerate the design of more rational targeted therapies. **DDW**

Dr Chris Torrance is a founder and Chief Scientific Officer of Horizon Discovery Ltd. He has significant oncology research and development experience, his principle expertise lying in cancer cell biology and drug discovery. In this field, he has led project teams taking drug targets from inception through to hit identification, lead optimisation and into preclinical studies. Chris holds a PhD from East Carolina University, USA and completed a Post-Doctoral position in the oncology lab of Professor Bert Vogelstein (Johns Hopkins University) where he pioneered the use of 'isogenic' X-MAN™ cancer model cell lines in HTS and drug discovery.

Eric Rhodes is Chief Technology Officer at Horizon Discovery. He joined Horizon at the start of 2012, bringing more than 15 years of gene regulation and gene editing experience. Prior to this Eric served as Director of Business Development at Sigma-Aldrich Corp where he helped Sigma launch its nuclease-based gene editing platform. Eric also served for 10 years as Vice-President of Business Development and Alliance Management at Sangamo BioSciences where he was responsible for putting target validation and technology licensing deals in place with more than 25 of the top pharma and biotech companies in North America, Europe and Asia. He has a degree in Microbiology and Immunology from UC Berkeley.