Tech News

The New Genetic Engineering Toolbox

By any account, the start of 2013 was extraordinary for would-be genome engineers. Between January 3 and January 29 six papers appeared in the pages of Science, Nature Biotechnology, and eLife, all leveraging a curious form of bacterial immunity called CRISPR/Cas to nip and tuck the genomes of human and mouse cells, bacteria, and zebrafish.

CRISPR/Cas is as essentially a bacterial adaptive immune system enabling cells to “remember” foreign nucleic acid invaders and silence them by site-specific DNA cleavage. Jennifer Doudna, a Howard Hughes Medical Institute Investigator and Professor of Biochemistry and Molecular Biology at the University of California, Berkeley, with Emmanuelle Charpentier at the University of Umeå in Sweden, unraveled the biochemistry of the process in late June 2012 and showed the system could also be reprogrammed to recognize different targets in vitro.

Clearly, the genome editing community was watching: Five other groups (plus Doudna’s) quickly picked up the ball and ran with it, demonstrating that CRISPR/Cas systems could create targeted double-stranded breaks in genomic DNA in live cells, the first step in genome editing. Manuscripts rolled into Science’s editorial offices a scant three months later.

“There’s a real hunger in the scientific community for a tool like this,” Doudna concludes.

Building on cellular tools

To appreciate the advance CRISPR/Cas represents, it helps to have some historical perspective. In the biological pleistocene—that is, prior to the mid-1990s—few options existed to manipulate the genome of a cell or organism, and those that did involved assembling a sequence in vitro, introducing it into cells, and hoping for homologous recombination to occur. There was no way to direct the process really, because there was no way to force the cell to take up the sequence of interest and copy it into its genome.

Today, that limitation is largely ancient history thanks first to custom DNA endonucleases (zinc-finger nucleases, TALENs, and meganucleases) and, more recently, CRISPR/Cas. As Dan Voytas, Professor and Director of the Center for Genome Engineering at the University of Minnesota, explains, these systems all work their magic by tapping into the same fundamental molecular process.

“The bottom line is that they are all ways to create a targeted break in a chromosome in living cells,” Voytas says. “That seems like a strange way to begin your precise modification, but once the chromosome is broken, [the cell] seeks to repair the break, and then we can direct how that break gets repaired.”

The cell uses—and genome engineers exploit—two primary mechanisms to repair double-strand breaks. The simplest and most efficient strategy is non-homologous end-joining (NHEJ), in which two DNA ends are simply glued back together. If that process worked perfectly, there would be little value in NHEJ as a genome-editing tool, but as it turns out, the process is error-prone. Bases are often added or deleted to produce small “indels” and, likely, either frame-shift or nonsense mutations within the coding sequence of a gene. The result: a quick-and-dirty gene knockout, no embryonic stem cells required.

“I call that the ‘laser method’,” says Stephen Ekker, Professor of Biochemistry and Molecular Biology at the Mayo Clinic Cancer Center, who uses custom endonucleases to manipulate zebrafish genomes. “We’re using a guided laser where you zap the DNA and then, poof! Some changes happen right there.”

The second repair strategy is homology-directed repair, the equivalent of the cell going to its backup hard drive to restore a damaged file. Normally, that correct information is contained on the second chromosome copy. But genome engineers can hijack that process too if they supply “donor” copies of their own containing anything from a point mutation or two to an entirely new gene sequence.

Designing custom endonucleases

So, how does one generate the initiating DNA lesion that induces these mechanisms? Until earlier this year, the only option was to build a custom restriction enzyme.

Researchers at Johns Hopkins University first developed this idea in the mid-1990s by coupling the endonuclease domain of the FokI restriction enzyme to the DNA-binding domain of zinc-finger transcription factors, creating what have come to be called zinc finger nucleases (ZFNs).

It turns out that zinc-finger transcription factors recognize DNA in a very modular fashion with each “finger” binding a specific and predictable trinucleotide sequence. This means it should be possible to target any desired sequence by simply assembling the proper set of fingers, three bases at a time, and expressing the resulting protein in cells.
The key words, though, are “should be”. For some researchers, the technology has been transformative. Model organisms that were refractory to basic genetic manipulations, including most vertebrates other than the mouse, were suddenly open to genetic manipulation. Sangamo BioSciences, a company that develops zinc-finger-protein-based therapeutics, has initiated human clinical trials with ZFNs targeting the cell surface receptor CCR5 to stymie HIV-1 infection (CCR5 is an HIV-1 co-receptor). And Sigma-Aldrich (which sublicensed the technology from Sangamo for research applications) has built libraries of ZFNs targeting every human, mouse, or rat gene ($4,000 each, according to the company’s website), or researchers can order custom ZFN for about $6,000 apiece.

But for other researchers, ZFNs have proved troublesome, being both expensive and temperamental. Voytas has a great deal of experience with ZFNs, having published some 20 papers on them and even co-founding the Zinc Finger Consortium to make the technology available to others. Yet, he says, “It’s a little bit hard to predict how the zinc fingers are going to work in any given array. You can’t just say, ‘I recognize GAT and it will always recognize GAT in every context I place it.’”

In 2009, two groups independently discovered another useful class of DNA-binding proteins in the pathogenic plant bacterium Xanthomonas. The transcription activator-like effectors (TALEs) also bind DNA with a modular and predictable architecture. But instead of gripping nucleic acids in groups of 3, TALEs bind DNA through a long series of 34-amino-acid repeats, each of which contains a central amino acid pair that specifies a single nucleic acid base in the recognition sequence. By assembling TALEs with ordered arrays of these repeats, researchers could specify their binding sites with nucleotide precision. In 2010 Voytas and colleagues at Iowa State University coupled a custom TALE with the FokI nuclease domain to create a so-called TALE nuclease (TALEN).

Today Voytas mostly uses TALENs in his plant research lab. (He is also Chief Scientific Officer for Cellectis Plant Sciences, a company that develops TALENs for agbio applications.) In fact, they have become the tool of choice for many in the genome-editing community. Ekker, for instance, has been using them to toy with the zebrafish genome. “I have no zinc finger that competes with what are some of our worst TALENs,” he says. “And our best TALENs, there’s no zinc finger or CRISPR system that competes.”

Ekker led one of two research teams that independently reported using TALENs to, for the first time, use an endonuclease to introduce novel DNA into the zebrafish genome by homology-directed repair. He and his colleagues used a novel high-efficiency TALEN system called GoldyTALEN, plus short single-strand oligonucleotides, to stitch in either a custom EcoRV or loxP site at specified locations within the zebrafish genome, in some cases with biallelic efficiency. “To our knowledge, these results represent the first description of successful [homology-directed repair, HDR] in zebrafish and the first demonstration of HDR using ssDNA as a donor template in vivo,” the authors write. (1) The second study, from Shuo Lin’s group at the University of California, Los Angeles, used long double-stranded donor sequences to introduce GFP into three specific zebrafish genes in vivo, albeit at lower efficiency.

Their popularity notwithstanding, developing TALENs is not exactly a walk in the park. They are long and highly repetitive and their assembly can be challenging, says J. Keith Joung, an Associate Professor of Pathology at Harvard Medical School, who uses TALENs for zebrafish and human cell-based work. Joung’s lab has developed three TALEN-design systems, including FLASH (Fast Ligation-based Automatable Solid-phase High-throughput system), a solid-phase method that uses libraries of preassembled cassettes of up to four TALE repeats to assemble the final molecule. Using this strategy, Joung’s team rapidly built 48 TALEN pairs targeting EGFP and 96 pairs targeting endogenous human genes. Every one of the EGFP-targeted TALENs worked in human cells, as did 84 of 96 of the endogenous gene-targeting molecules. “We effectively don’t think there are any real targeting range limitations to the TALEN platform,” he says.

An RNA alternative

What limitations, if any, will constrain the CRISPR/Cas system remain to be determined.

Following infection by exogenous DNA, some bacteria and archaea use the CRISPR/Cas system (short for Clustered Regularly Interspersed Short Palindromic Repeats [CRISPR]/CRISPR-associated [Cas]) to copy a segment of the invading sequence into a specific region of their chromosomes, thereby archiving knowledge of the infection for the cell and all its progeny. Upon subsequent invasion with a new bit of foreign nucleic acid, the cell can refer back to that archive to see if the infection is one it’s seen before, at which point the cell destroys the incoming DNA. “These repetitive loci serve as molecular ‘vaccination cards’ by maintaining a genetic record of prior encounters with foreign transgressors,” Doudna

Keith Joung and his team recently used the CRISPR/Cas system to generate targeted mutations in zebrafish.
wrote in a review appearing months before the seminal 2012 Science paper (2).

Led by Doudna’s postdoc Martin Jinek and researchers in Charpentier’s lab, that paper demonstrated that Strep- tococcus pyogenes protein Cas9 is an RNA-guided DNA endonuclease, and furthermore, that Cas takes its marching orders from a short guide RNA molecule called a crRNA, plus a second required (but non-targeting) RNA called tracrRNA. The team then showed how the system could be modified: First, they combined the crRNA and tracrRNA into a single chimeric molecule, thereby making the method more user-friendly; then, by supplying a custom crRNA, they targeted the system to other DNAs, in this case, to a plasmid-bound copy of GFP in vitro.

Within months, teams led by Doudna, Joung, Harvard geneticist George Church, Feng Zhang at the Broad Institute of MIT and Harvard, Jin-Soo Kim at Seoul National University in South Korea, and Luciano Marraffini at the Rockefeller University, reported adapting Charpentier and Doudna’s modified system for genome editing in cultured cells or in vivo.

According to Doudna, the system is fast, easily programmed, inexpensive, and can target essentially any 20-base sequence so long as it is immediately followed by a 3’-NGG (the so-called “protospacer-adjacent motif”). “This is probably why the six papers came out as quickly as they did,” she says. “[The system] didn’t require a lot of additional fiddling or optimization.”

Church applied bacterial CRISPR/Cas to human cells in culture, using the system and homologous recombination to restore activity to an inactivated genome-integrated GFP gene and to insert a selectable market at an “endogenous” viral locus on chromosome 19. In one experiment in HEK 293T cells, the team directly compared TALENs with CRISPR/Cas9, finding that the RNA-based system was about 20 times more efficient at inducing homology-directed repair (8% efficiency vs 0.37%). CRISPR/Cas-directed NHEJ was up to 38% efficient in K562 cells, but much less so in induced pluripotent stem cells (<4%).

Naturally, Church is trying to optimize the process. But already he imagines new applications in human gene therapy, for instance to change multiple copies of specific histocompatibility genes. “If [the process] really is 20 times better than zinc fingers and TALENs, then some human gene therapies that are hurting for extra efficiency could now be conceivable,” he says.

Church’s team also demonstrated one feature of the CRISPR system that isn’t nearly as easy with ZFNs or TALENs: multiplexing. Because the system relies on a single nuclease guided by short guide RNAs, CRISPR/Cas can theoretically generate multiple genome edits simultaneously simply by adding multiple guide RNAs. In this case, the team showed that introduction of two guide RNAs resulted in “high-efficiency” deletion of the intervening sequence, “demonstrating that multiplexed editing of genomic loci is feasible using this approach.” (3)

Not to be outdone, Joung’s team took CRISPR/Cas in vivo, targeting 10 zebrafish genes by NHEJ in embryos. Though 2 of the targeting reactions failed, the remaining 8 produced mutations at rates of from 24% to 59%.
Targeting the future

One unresolved question for all genome-editing systems, especially CRISPR/Cas, is what happens beyond the targeted sequence—so called off-target effects. “There are still open questions about specificity on all these different platforms,” Joung says.

In 2011, Joung coauthored one of two studies examining ZFN specificity. Though both studies used the same ZFNs and the same target (CCR5), the resulting lists of off-targets did not overlap, says Joung, “Which implies to me that neither one is capturing the full universe of off-targets, because they’re missing the ones found by the other method.”

The obvious approach to assessing off-target effects is simply sequencing the modified genomes. That’s precisely what Chad Cowan at the Harvard Stem Cell Institute, and colleagues did after using TALENs to generate mutations in human induced pluripotent stem cells. Sequencing six exomes, they identified two indels other than the expected ones, but lack of homology to targeted sequences suggests these sites are unrelated to the TALEN, the team says. The team also performed whole transcriptome sequencing of these cells failed to identify any genes, other than the specific target, whose expression was downregulated. “At least in bacteria, this is an incredibly sequence-specific process,” Doudna says.

And it is one that will likely develop over the next few months, she adds, which should definitely help to define the system’s potential and limitations. “It’s a really great example of how fundamental basic research, which was not aimed at any particular target or goal or certainly a particular application, led to the discovery of a system that may turn out to be a really transformative technology for genome engineering.”

Besides Doudna and Lim’s work on a Cas-based repressor, Joung has coupled TALE binding domains to transcriptional activation domains to create highly active “TALE activators.” Robert McKnight at the University of Utah is building TALE methyltransferase fusions (TALEMs) to manipulate the epigenome one CpG at a time, and imagines other applications to probe hydroxymethylation, histone modification, and more. “The potential is tremendous,” he says.

Fortunately, researchers who wish to tap into that potential needn’t reinvent the wheel. Voytas, Joung, and Doudna, among others, have made their TALEN and CRISPR/Cas reagents available at AddGene. And Ekker’s lab recently launched an online TALEN-design tool called Mojo Hand (www.talen-design.org). Looks like its time to clear some space in your toolbox.