CRISPR/Cas9: Tools and Applications for Eukaryotic Genome Editing

Fei Ann Ran
Broad Institute
Cambridge, Massachusetts

ran@fas.harvard.edu

I will provide some background on the CRISPR/Cas9 technology, some of the rationale for how we came to develop and use this tool, and I will address immediate questions concerning the specificity of the technology. I will also discuss some of the more interesting applications.

Figure 1 reflects how the cost of DNA sequencing has decreased dramatically over the past two decades due to technological progress. As a result, there has been an explosion of data, not only in the sequences of different species, but in sequence differences between individuals within species, between cell types and between diseased and healthy cells. It suffices to say that this is an exciting time to be working in the field of genome engineering.

Genome Engineering

Typically, genome engineering is achieved by leveraging the cell's own repair machinery. This can come from the error-prone NHEJ pathway that leads to insertion/deletion (indel) mutations, which can be used to knock out genes, or, alternatively, we can supply a repair template to overwrite the site of a double-stranded break (DSB) for more-precise genome engineering via the HDR pathway (Figure 2).

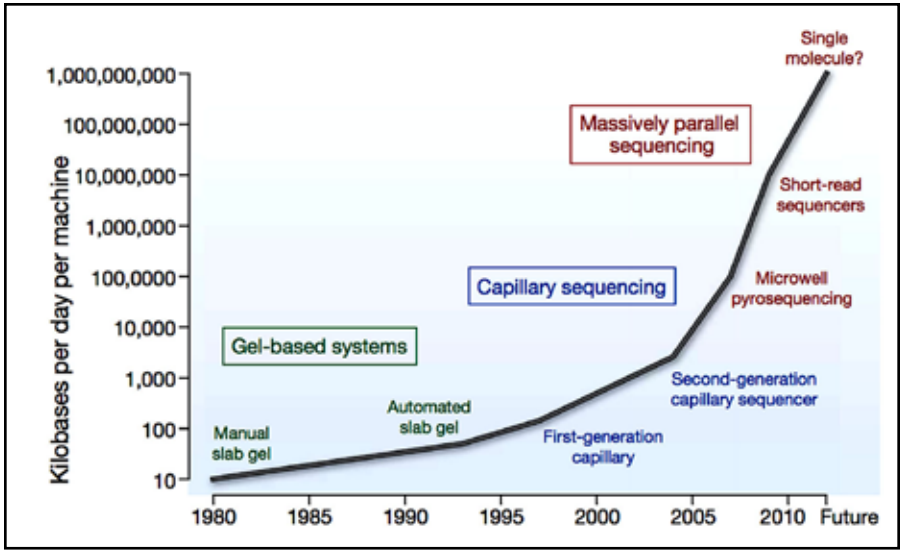

Figure 1. Advances in DNA-sequencing technologies.
(Stratton MR *et al.*, 2009)

When we started working on CRISPR/Cas technology[1], several well developed tools were already being used—and still are being used—to achieve impressive results in biotechnology, medicine, agriculture, and other fields. At the outset, we were interested in developing an alternative technology to make cloning easier at lower cost with greater scalability.

The CRISPR locus, including the hallmark repetitive patterns of crRNA, was discovered in the genome of *Escherichia coli* over 25 years ago, and, at the time, no one knew what they were. In the 2000s, it was discovered that this was a defense system against viral infection. Figure 3 illustrates a phage injecting its genome into a bacterium; a portion of the viral genome is inserted by the bacterium into its own genome. These inserts were initially called spacers, which are vitally important for CRISPR-system function. One of the requirements for what sequence can be incorporated and inserted into these CRISPR-loci in bacteria is what's called a protospacer adjacent motif (PAM). This is important because a given species may have one or multiple types of CRISPR systems, and each CRISPR system may have a unique PAM.

The CRISPR system that we started working with—now one of the most widely adopted—comes from *Streptococcus pyogenes* (Figure 4) and the PAM for that species is

[1]Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins are found in many bacteria and most archaea. CRISPR-Cas systems use sequences derived from plasmids and phages to activate Cas endonucleases to neutralize those plasmids and phages via RNA-guided sequence-specific DNA cleavage, thus blocking their transmission and creating simple acquired immunity.

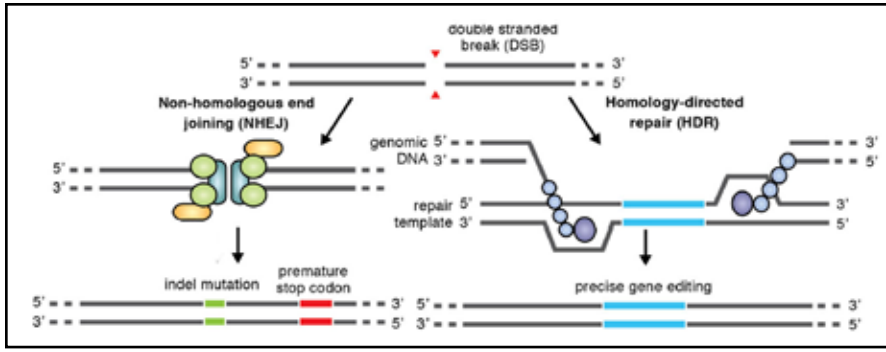

Figure 2. DNA double-stranded breaks facilitate alteration of the genome. (Ran *et al.*, 2013)

an NGG-trinucleotide motif, which means that the *S. pyogenes* CRISPR system incorporates only sequences adjacent to NGG. After integration, the bacterium can carry out the "execution" part of its defense. During this stage, there are several key players. One is the CRISPR array that becomes transcribed as a long precursor CRISPR RNA (pre-crRNA). This is a string of direct repeats flanked by spacers and this array can go on for 30 to 60 different spacers, as a single long transcript. In the presence of the trans-activating CRISPR RNA (tracrRNA) and the Cas9 nuclease, the pre-crRNA:tracrRNA duplex gets processed to its mature form, consisting of single units of processed spacers and direct repeats hybridized to the processed tracrRNA. Now the mature crRNA:tracrRNA duplex can guide the Cas9 to target any sequence that is complementary to the spacer. The protospacer adjacent motif (PAM) is again crucial for DNA cleavage by Cas9. Cas9 will cleave only targets that are immediately adjacent to a PAM.

When we started working on this, the tracrRNA hadn't been discovered yet. This discovery came from Emmanuelle Charpentier's lab, and helped kicked off genome engineering using CRISPR/Cas9. At that time, two other developments also emerged (Figure 5): (1)

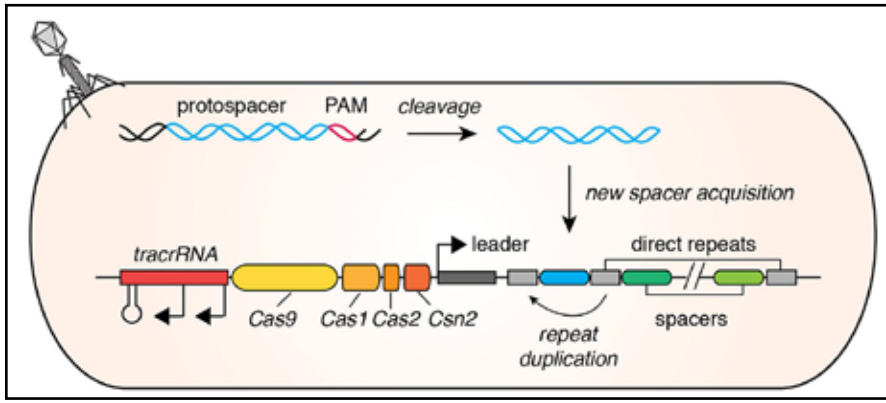

Figure 3. Clustered regularly interspaced palindromic repeats (CRISPRs).

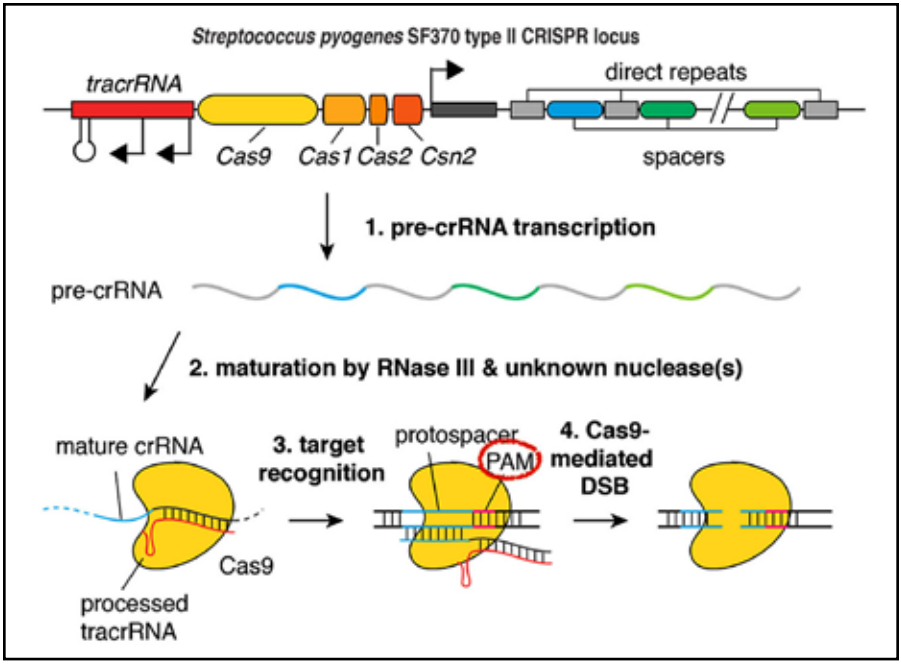

Figure 4. *Streptococcus pyogenes* CRISPR system.
(Cong *et al.*, 2013)

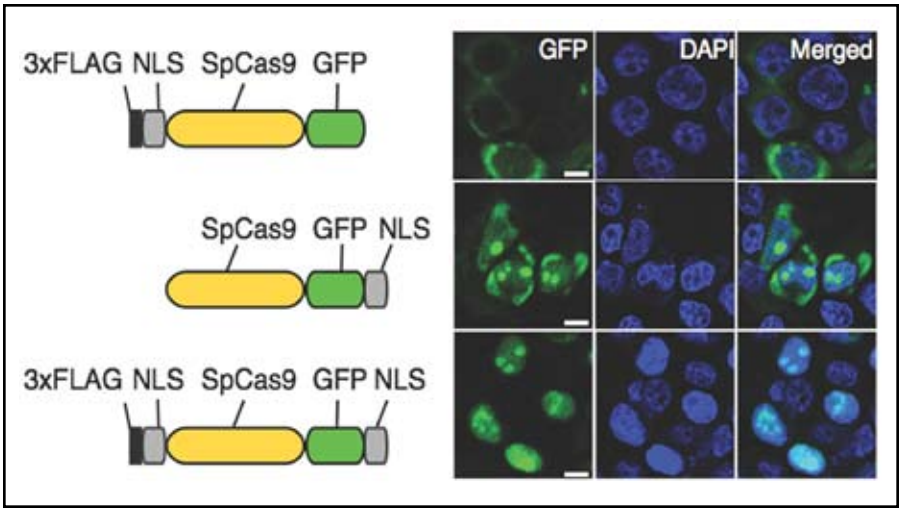

Figure 5. Adapting *S. pyogenes* Cas9 (SpCas9) for eukaryotic expression[2].
(Cong *et al.*, 2013)

[2]DAPI=4',6-diamidino-2-phenylindole, a fluorescent stain that binds strongly to A–T-rich regions in DNA.

you can fuse the spacer and the repeat and the tracrRNA into a single chimeric RNA; and (2) you can use a single chimeric RNA and Cas9 to program the cleavage of DNA targets in an *in vitro* cell-free lysis reaction.

We built upon these exciting discoveries, but at the same time, nobody knew if this was going to work in mammalian cells. We modified two systems to get this working robustly in eukaryotic systems. One issue was that, obviously, bacteria don't have nuclei, whereas mammalian and other eukaryotic cells do, and so we tagged NLS (nuclear localization signal) sequences to Cas9 and also codon-optimized it for better eukaryotic expression. By doing this, we were successful in moving the Cas9 enzyme into mammalian nuclei. These experiments were done in human embryonic kidney (HEK) cells (Figure 5).

We started these experiments with the same type of chimeric RNA as described earlier. But we weren't having luck targeting every locus. So, we went back to optimize the RNA components and extended the tracrRNA portion of the chimeric RNA to its original full length that is expressed by the bacteria. We call this single-guide RNA (sgRNA) (Figure 6). The sgRNA has an invariant scaffold region and a spacer region or the guide proper region that base pairs with the target and once it brings Cas9 to the locus of interest, Cas9 makes a double-stranded break about 3–4 base pairs upstream of PAM.

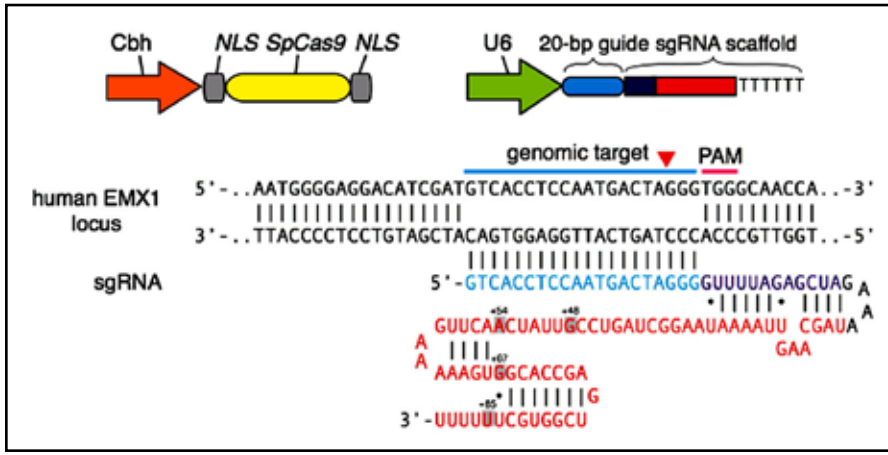

Figure 6. New single-guide RNA (sgRNA) design improves cleavage efficiency. (Hsu *et al.*, 2013)

With these two modifications, we were able to increase the efficiency of Cas9-mediated genome engineering in mammalian cells. Figure 7 shows an enzymatic assay, SURVEYOR, which we use to measure the efficiency of genome editing and—without going into detail—the numbers on the bottom are the percentages of transfected cell populations that have acquired indel mutations. This is an example of the more error-prone NHEJ way, which can be leveraged to do simple gene knockouts by creating mutations in the coding region of the gene. In the absence of any type of selection, with a transient transfection we see very high modification rates in mammalian cells.

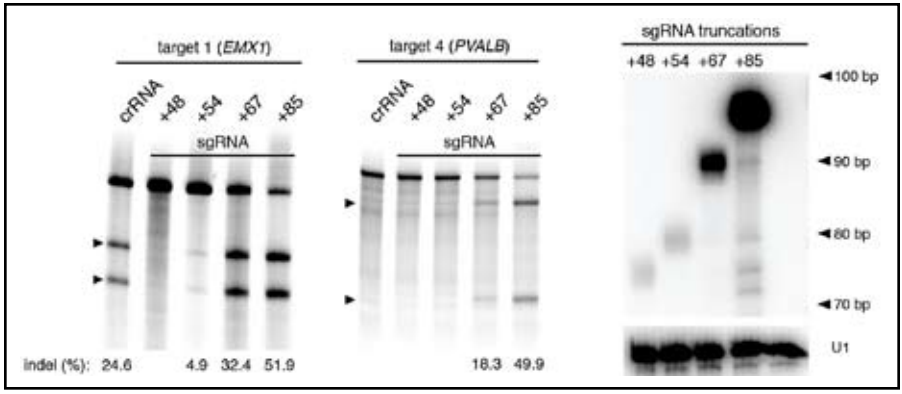

Figure 7. Extended single guide RNA (sgRNA) improves cleavage efficiency.
(Hsu *et al.*, 2013)

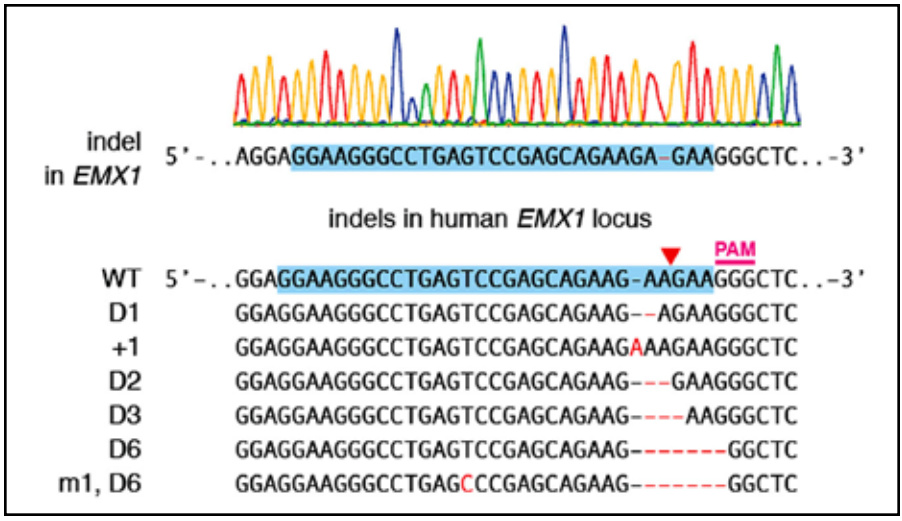

Figure 8. Cas9-mediated indel mutations.
(Cong *et al.*, 2013)

Figure 8 shows what the indel mutations look like when sequenced. Again, most are centered about 3–4 bases upstream of PAM. Cas9 was a very easy-to-engineer technology, because all that was necessary to target a locus was to provide an RNA template to Cas9. We thought we could multiplex the system, in other words knock out multiple genes in the same cell. Initially, we tried knocking out just two genes in the cell, which involved co-delivering two guides in Cas9. Again in a transient transfection of mammalian cells, both genes underwent fairly significant levels of indel modification. Since then other people have iterated those to a much higher order of multiplexing.

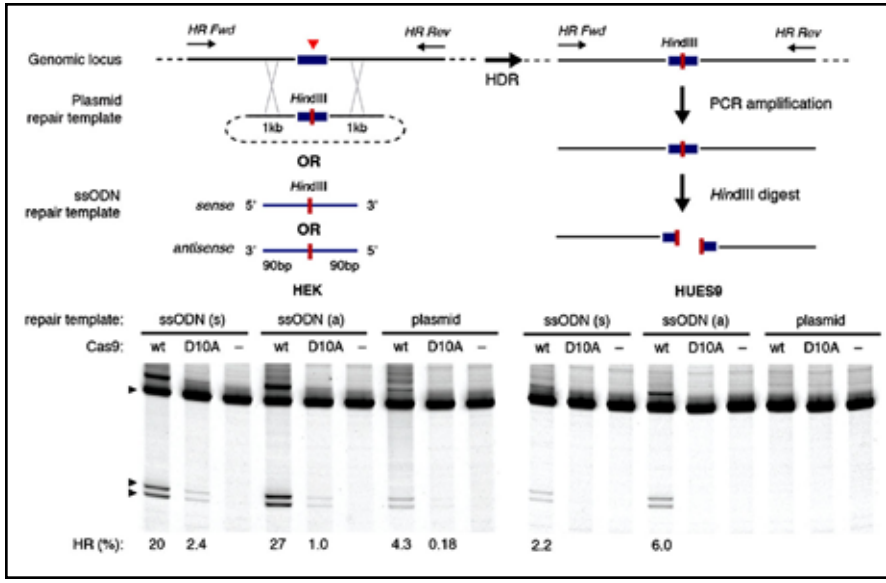

Figure 9. Using CRISPR-Cas9 to mediate precise gene editing.
(Ran *et al.*, 2013)

Similar to zinc fingers and TALENs, Cas9 is also a mediator of HDR through creating double-stranded breaks. Figure 9 provides an example of using Cas9 for introducing a pair of restriction sites into the genome through HDR. One can introduce restriction sites, epitope tags, or SNPs into a locus of interest by building a traditional homologous repair template in the form of a plasmid with 1–3-base flanking arms, or one can use a single-stranded DNA oligo to repair the template to introduce these types of small changes. In the absence of any selection, we see again fairly high levels of HDR being mediated by Cas9, and one can titrate these numbers with additional screening or selection.

By way of a quick summary, I hope I have made a convincing case that Cas9 is an easy-to-use system for both introducing indels as well as mediating HDR. Recently, the crystal structure of Cas9, alone or in complex with the guide RNA and target DNA, was solved by three groups. Figure 10 shows that the enzyme has a bi-lobed structure. At the top is a domain mostly responsible for recognizing guide RNA and target DNA, and at the bottom of the Cas9 enzyme are the nuclease domains. These domains create a positively charged groove where DNA and RNA sit.

We created a pipeline for rapid generation of cell-line models in a span of about a month from the *in vitro* design of sgRNAs—which we have a website tool to help—to reagent construction and functional validation and expansion of cell lines (Figure 11). We published this in *Nature Protocols* in 2013 and have deposited Cas9, and GFP, puro, and nickase versions, at Addgene.

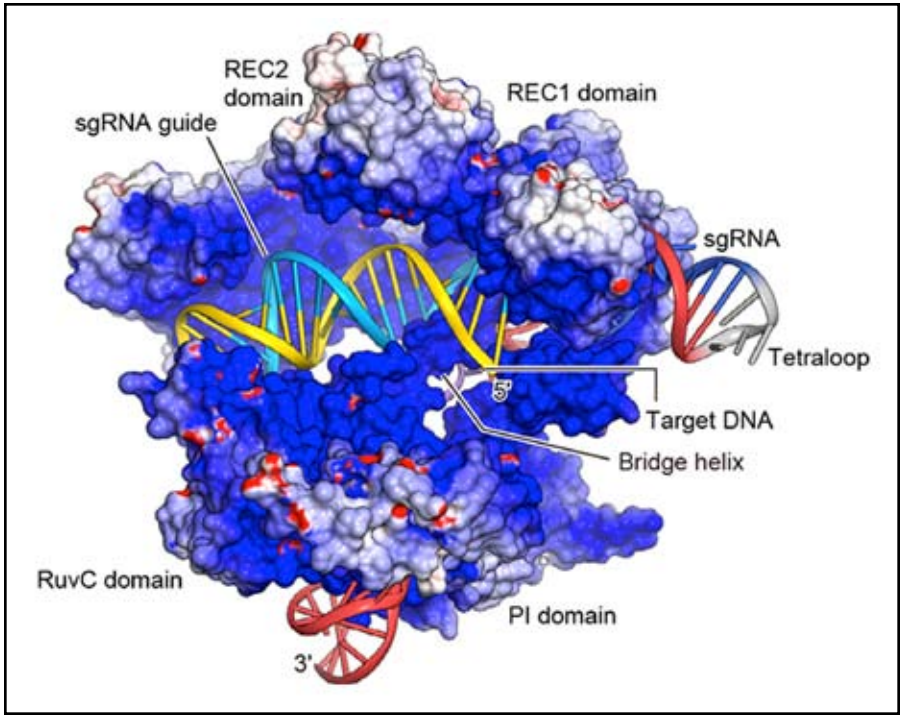

Figure 10. A new system for efficient mammalian genome cleavage.
(Anders *et al.*, 2014; Jinek *et al.*, 2014; Nishimasu *et al.* 2014)

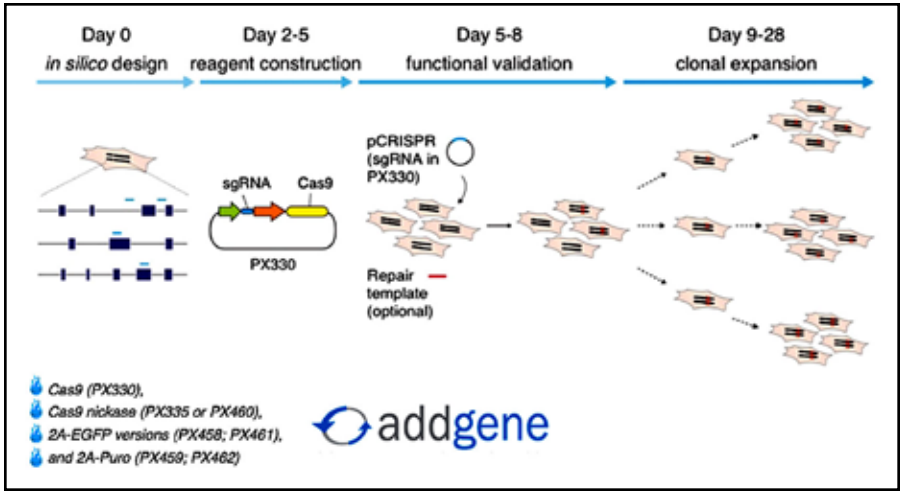

Figure 11. Pipeline for rapid generation of cell-line models.
(Ran *et al.*, 2013)

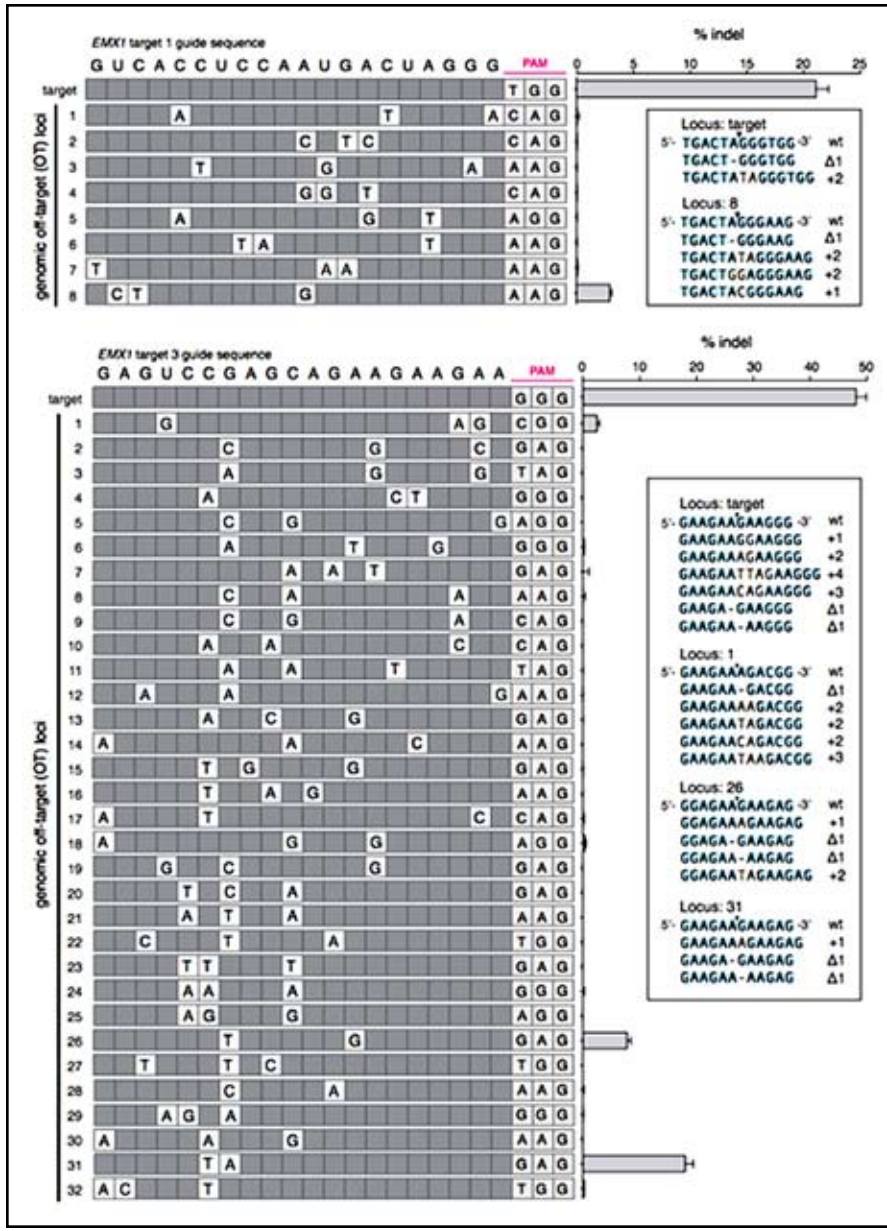

Figure 12. Cas9 can have off-target effects.
(Hsu *et al.*, 2013)

Knowing now that Cas9 works well, one of the immediate concerns in everyone's mind was the specificity of the system. Figure 12 shows that Cas9 does have off-target activities and these tend to occur when there are mismatches between the guide RNA and the target DNA on the PAM-distal side of Cas9. Several groups have come up with independent ways of improving the specificity. One idea is that you can actually truncate the guide sequence, which is usually 20 nucleotides long. Shortening this to a 17-, 18- or 19-nucleotide sequence is sufficient to improve the specificity of Cas9 by a significant amount. Another idea is to use an enzymatically dead version of Cas9 and tow around a Fok1 nuclease and rely on the obligate heterodimeric properties of Fok1 to increase that specificity.

The tack that our group took—which was introduced by Dana Carroll[3]—was using Cas9 as a double nickase. If you situate two units of Cas9 nickase on opposite strands of DNA, then a nick plus another nick equals a double-stranded break. This works efficiently and it works across a wide number of distances. Double nicking can happen as far as 100 nucleotides away from each other.

Returning to specificity, Figure 13 shows that Cas9 nickase can increase specificity of the system by several orders of magnitude. One of the other advantages of using double nicking is that it creates staggered cuts, which is reminiscent of cloning using restriction enzymes. It turns out that we can do something similar in cells, and if you have a repair template or insertion template that has corresponding arms that can be inserted directly into the staggered cuts, then we can essentially do ligation-based cut and paste of the template directly into cells. So, this is another alternative strategy to HDR or indel for specific genome editing.

To sum that part up, for specificity considerations what one would like to do ideally is select unique parts of the genome to target and avoid sites with large numbers of off-target matches in what is considered a seed region or the PAM-proximal region of sgRNA. Also, one can use techniques such as paired sgRNAs for double nicking or shorter truncated sgRNA guides, or both together, to improve the specificity of the system. And to improve the activity of the system, the guide should always begin with a G and avoid poly-T tracts to prevent premature transcriptional termination. Taking these together, one can design very efficient and specific guides.

Finally, I will talk briefly about applying Cas9 towards *in vivo* cell editing. One of the main challenges to using Cas9 in adult somatic tissues is delivery. Currently, one of the most clinically promising vehicles is adeno-associated virus (AAV), which is used for several human clinical trials and AAV1 was approved as a gene therapy vehicle in Europe. For AAV delivery, the *S. pyogenes* Cas9 (SpCas9) that everybody has been using so far is a little too big to fit with its sgRNA and all the regulatory elements in a single vector. So our lab has developed an additional Cas9 from *Staphylococcus aureus* (SaCas9), which is small enough to be squeezed into a single AAV vector along with its sgRNA. SaCas9 has a different protospacer adjacent motif (PAM), NNGRRT, which is relatively permissive, as it's required to be present next to the target for Cas9 binding and cleavage. We were

[3]Pages 25–27.

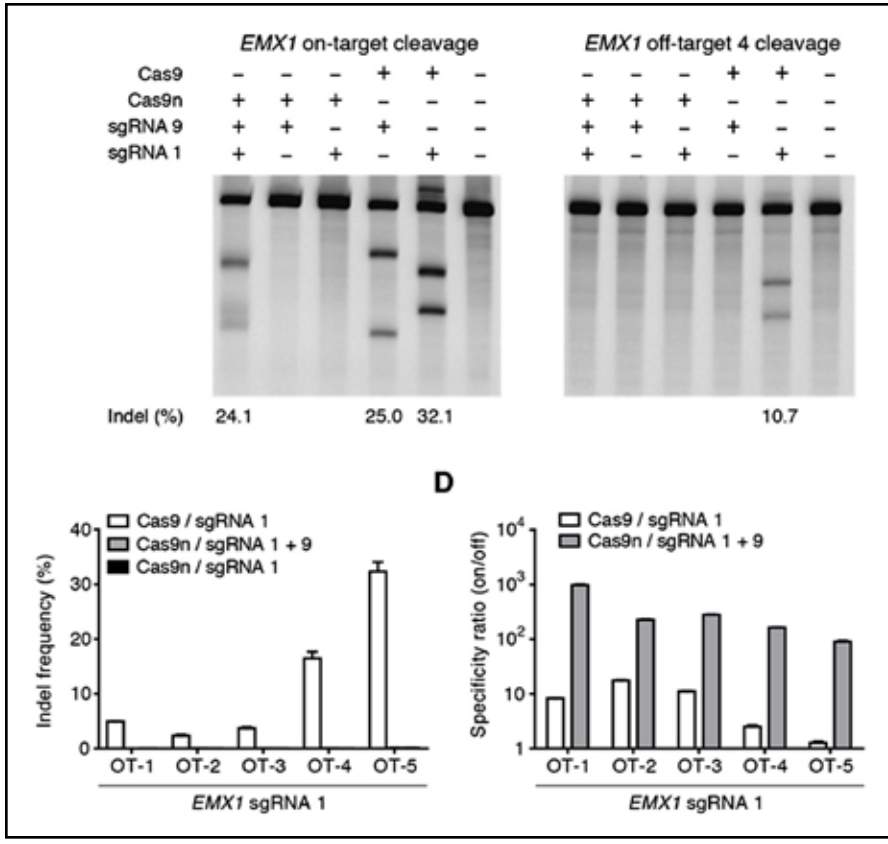

Figure 13. Double nicking improves specificity.
(Ran *et al.*, 2013a)

able to package SaCas9 and its sgRNA into AAV with the AAV8 capsid and inject the particles into animals via the tail vein. Depending on the serotype of the AAV, other tissue types of interest can be targeted.

Figure 14 shows some of our preliminary data. We targeted the ApoB gene in the mouse liver; ApoB knockout leads to an oil-droplet-accumulation phenotype that we can observe. Next, we tried a promising target for treatment of hypercholesterolemia, *Pcsk9*, which regulates the cycling of LDL-receptors. One week after injecting AAV bearing SaCas9 and sgRNA against *Pcsk9* into the animals, we saw a 40% gene modification in the liver and a depletion of serum Pcsk9 levels in the treated animals. This is a work in progress, but we are excited about the possibility of expanding the use of Cas9 towards DNA-editing in somatic tissues of adult animals.

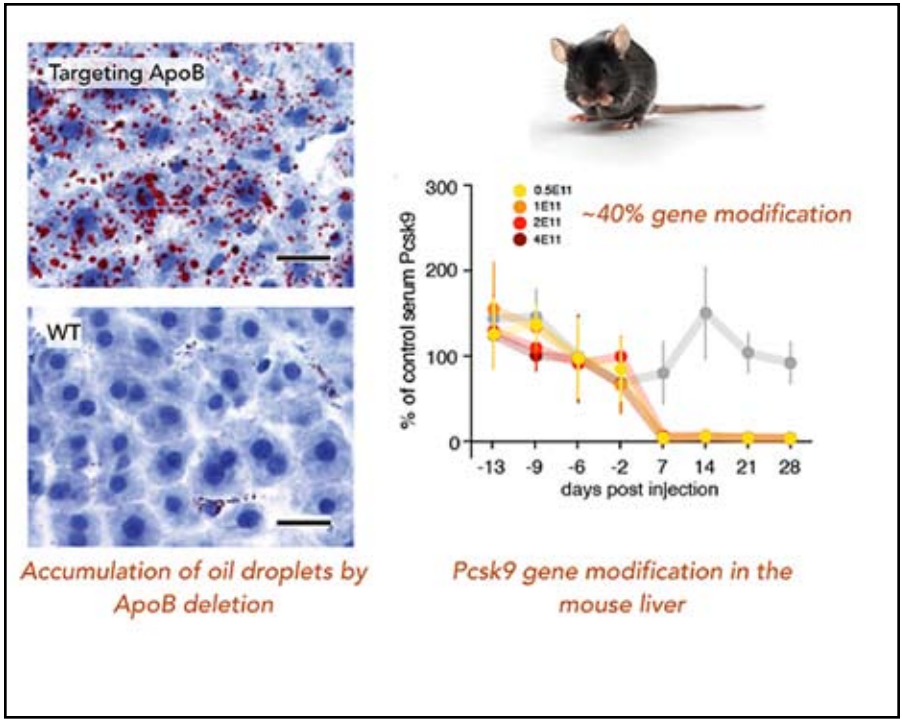

Figure 14. Sa Cas9 can be delivered by AAV to target genes *in vivo*.

References

Anders C *et al.* (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513 569–573. DOI:10.1038/nature13579.

Cong L *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339 (6121). DOI:/10.1126/science.1231143.

Gasiunas G *et al.* (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences of the USA 109(39) 15539–15540. DOI: 10.1073/pnas.1208507109.

Hsu PD *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology 31 827–832. DOI:10.1038/nbt.2647.

Jinek M *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 17 337(6096) 816–821. DOI: 10.1126/science.1225829.

Jinek M *et al.* (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343(6176). *DOI:* 10.1126/science.1247997.

Nishimasu H *et al.* (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156 935–949.

Ran FA *et al*. (2013) Genome engineering using the CRISPR-Cas9 system. Nature Protocols 8 2281–2308. DOI:10.1038/nprot.2013.143.
Ran FA *et al*. (2013a) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154(6) 1380–1389. DOI: 10.1016/j.cell.2013.08.021.

ANN RAN is a postdoctoral fellow in Feng Zhang's laboratory at the Broad Institute of MIT and Harvard and a junior fellow at the Harvard Society of Fellows. She graduated *cum laude* with a BS from Yale in 2006, and received her PhD in biochemistry from Harvard in 2014. In the Zhang lab, she co-developed Cas9 for mammalian genome engineering and established methods for rapid generation of cell lines using Cas9, with reagents now widely distributed by Addgene.

DR. RAN has presented her work on improving Cas9 specificity using the double-nicking method and elucidating Cas9 and guide-RNA structure-function relationships at a number of conferences. She received the Meselson Prize at Harvard for the "most beautiful experiment" in 2013 and was a finalist for the Regeneron Creative Innovations Prize.