

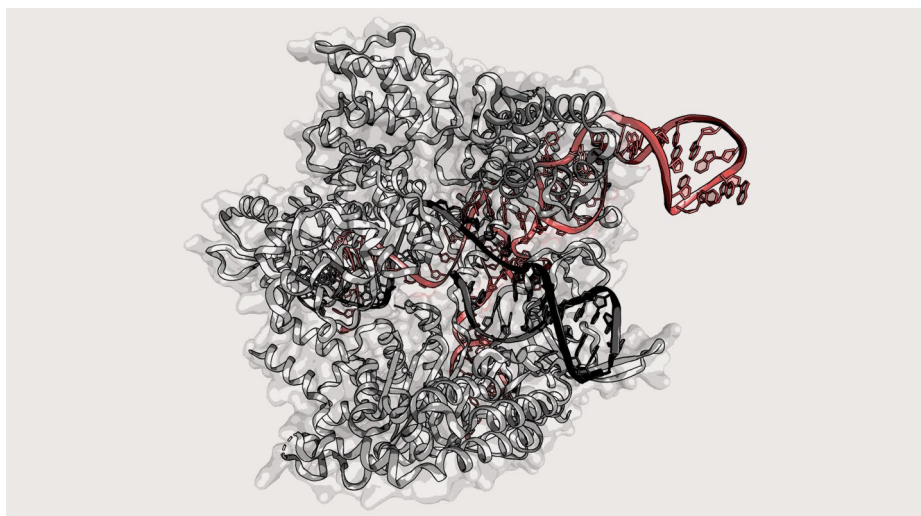
Guide RNAs: it's good to be choosy

CRISPR, now a Nobel-Prize-winning technology, comes in many flavors. And so do gRNAs.

Vivien Marx

“Nom nom nom. Me love gRNAs!” says Cas M(onster). That’s the caption to a cartoon by science illustrator Maya Peters Kostman of Cookie Monster munching on a cartoon genome. This guide RNA (gRNA) cartoon is part of a [blog post](#) with guide RNA design tips for CRISPR genome editing by John Doench, who directs research and development at the Broad Institute of MIT and Harvard’s Genetic Perturbation Platform. The post is on Addgene’s site, a resource for plasmid sharing. Beyond Cas M(onster) and Doench, others too may love — and certainly love to use — gRNAs, which are the targeting components of gene-editing methods that harness aspects of a bacterial immune system to make changes to DNA or RNA. New gene-editing techniques, such as prime editing, keep coming. “The field is already developing very fast, so it’s hard to imagine it going faster,” said Jennifer Doudna from the University of California Berkeley at a press conference. She is a 2020 Nobel laureate in chemistry for her CRISPR work. Says Hyongbum Henry Kim, a researcher at Yonsei University College of Medicine in Seoul, “I am excited to hear that Nobel Prize is given to the field of my research.” As the field basks in Nobel glory and gallops on, the gRNA family will continue to grow and diversify, and so will ways to assess gRNAs^{1–5}. There’s more to gRNA choice and use than the ‘Add to Cart’ click on an Addgene order. To help flatten the gRNA learning curve, methods developers have some recommendations.

The natural bacterial CRISPR system in *Streptococcus pyogenes* involves two RNAs, the crRNA and the tracrRNA, says Doench, which are fused together to build a single guide RNA (sgRNA). The Doudna and Charpentier labs⁶ showed DNA can be cleaved in a CRISPR-directed, programmable way, says Kim, and by combining a tracrRNA and a crRNA into an sgRNA, they simplified the components of genome editing. This simplification matters, he says, especially for downstream applications in the clinical realm, and it reduces effort, cost and time for labs. Like theater stagehands who get things to where they are needed, gRNAs recognize a sequence in a targeted genome where cutting or nicking enzymes get to work. Cas9 protein



gRNAs direct Cas9 to target sites in a genome where enzymes can get to work. Here a 3D structure of SpCas9, PDB ID 4UN3, with the gRNA (red). Credit: Kleinstiver lab, MGH.

directly recognizes GG dinucleotide short protospacer-adjacent motifs (PAMs), says Kim. For a while, there were limitations: if the target sequence did not have GG, or CC on the opposite strand, CRISPR could not be used at the target sequence. But several labs have developed Cas proteins that recognize a wider variety of PAM sequence. Researchers also have introduced modifications to gRNAs, such as ones that increase transcription and stabilize the structure, says Doench. For example, prime-editing guide RNAs (pegRNAs) are not fundamentally different, he says; they are gRNAs with added functionality.

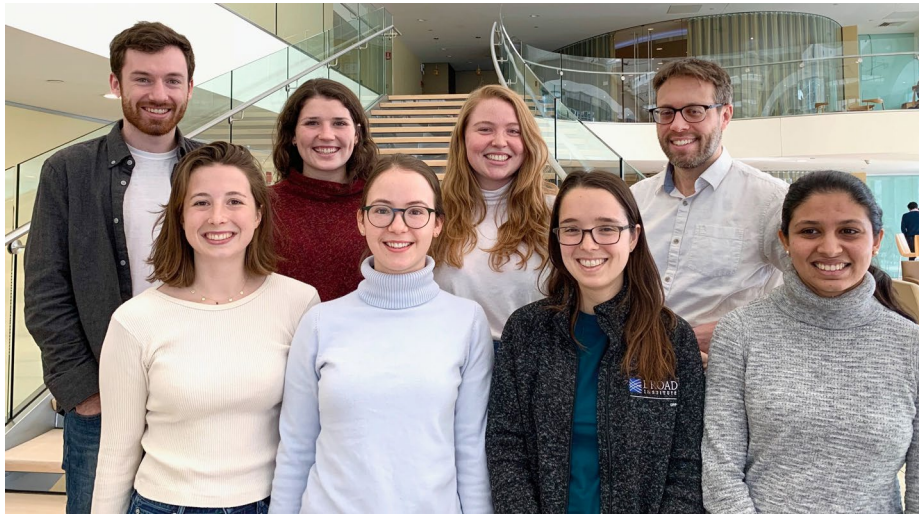
For Cas enzymes such as Cas12a, says Ben Kleinstiver, a researcher at Massachusetts General Hospital, the gRNA includes just the crRNA and not the tracrRNA. Kleinstiver, MGH’s Keith Joung and others recently engineered Cas12a variants to recognize more PAMs⁷. CRISPR–Cas enzymes have been adapted in a number of different ways, such as for base editing and prime editing, techniques that aim to nick one but not both DNA strands, says Kleinstiver, and gRNAs have been adapted to give them new functionality, too. For example, he says, some labs have tweaked the synthetic loop between crRNA and tracrRNA; RNA aptamers can be added to the 3’ end of a gRNA. These

aptamers can interact with RNA-binding proteins that can fuse to other heterologous proteins, which recruits them to the Cas9–gRNA complex. Other gRNA variants are truncated gRNAs with 14- or 15-base-pair spacers that can render a wild-type SpCas9, the most commonly used Cas enzyme, inactive; these might be used for gene editing or gene regulation applications, says Prashant Mali, a University of California San Diego researcher.

In silico choices

Guide RNA selection will vary by experiment type. Generally speaking, researchers will want to manually inspect their gRNA and target site, says Kleinstiver, and avoid polynucleotide or repetitive stretches. Poly(T) runs can act as transcriptional terminators, which can ruin production of a good gRNA.

Researchers can find many online gRNA design tools. In a recent overview⁵, Doench and his Broad colleague Ruth Hanna note at least 30 web-based gRNA design tools. More exist as downloadable tools, and companies offer additional ones. As Hanna and Doench assessed pros and cons, they looked, for example, at how the tools calculate on- and off-target scores and whether they support techniques such as CRISPR interference (CRISPRi), used to suppress transcription,



John Doench (top right) and his lab work on gene editing and the aspects that shape good gRNA behavior. Among the lab's tools is a 'cutting frequency determination' (CFD) score. Credit: O. Bare

or CRISPR activation (CRISPRa), in which transcription is activated. As each tool serves distinct purposes, labs might want to mix and match them. The gRNA design offerings get rather sparse when it comes to newer techniques such as prime editing. Kim says researchers should look for tools that they find easy to use and that are fairly accurate.

Doench and colleagues worked on predictive models and high-throughput assays to assess gRNA efficiency for editing mouse and human genomes. That includes ways to assess and predict on-target and off-target activity of the Cas9–sgRNA complex and the ways sgRNA interacts with DNA. The team teased out some guidelines for guide RNA design, which led to their Rule Set 1 and Rule Set 2 — quantitative models to assess sgRNA on-target specificity. The team's sgRNA libraries are found on Addgene. Jennifer Listgarten, then at Microsoft Research and now at the University of California Berkeley, and her Microsoft Research colleague Nicolò Fusi built the predictive models and talked plenty with Doench and colleagues. The on-target model is *Azimuth* and the off-target model is *Elevation*. At the time, recalls Listgarten, the team precomputed all off-target data across the whole genome for any guide that targets any human protein-coding gene. "This required some vast amount of compute and John didn't believe us when we said we could do it." The models are available [online](#) and built into the Broad's web application, *CRISPick*. Using the portal, scientists can apply the models, rule sets and scoring methods to pick and rank guide RNA sequences for a given target; minimize the probability of off-targets; and maximize on-target efficiency.

Colloquially, scientists sometimes mention the ranking and scoring approaches as 'Doench rules,' which amuses Doench. "I certainly never call them that," he says. "What I'm more proud of, to be honest, is that we came up with a 'cutting frequency determination' (CFD) score that has gotten some traction in other labs for looking at off-target effects," he says. This CFD score, a measure of likely off-target cutting by a given sgRNA–DNA interaction, is also part of the Broad's web resource. Among the other gRNA design sites that apply Rule Set 2 and CFD scoring are [GUIDES](#), for human and mouse genomes edited with the Cas enzyme SpCas9, and [CRISPick](#), for human, mouse and rat genomes and the Cas enzymes SpCas9, SaCas9, AsCas12a and enCas12a.

To knock out a gene, says Kleinstiver, there are many options for selecting a gRNA for putative target sites. Generally, his lab combines tools to prioritize gRNA options, including those that predict gRNA activity and editing outcomes, such as GPP sgRNA Designer, [CHOPCHOP](#), [CRISPOR](#), [inDelphi](#) and others. The team uses off-target nomination tools such as [CasOFFinder](#). Mali likes and uses the Broad's GPP sgRNA Designer. "I do also like the one from Raj Chari," says Mali, noting his bias because he collaborated with the Chari lab on the team's tool, sgRNA Scorer. Chari directs the genome modification core in the Laboratory Animal Sciences program at the Frederick National Laboratory for Cancer Research, which is affiliated with the US National Cancer Institute. There is [sgRNA Scorer 2.0](#) and also the dbGuide database, a compilation of validated guide RNAs sequences used in experiments involving human and mouse cells.

Commercial takes

Agilent had been working on RNA synthesis chemistry and then applied it to gRNAs. They found they could make longer, chemically modified guides to reduce off-target activity and maintain on-target activity, says Agilent's Laurakay Bruhn. She leads interdisciplinary research teams related to, among others, nucleic acid synthesis and gene editing. She and others worked with Matthew Porteus's lab at Stanford University and showed these modified gRNAs performed better than unmodified ones, especially in cells such as stem cells and T cells that are tricky to edit, she says. A gRNA can be modified to tweak its stability or make it able to carry a fluorescent label to a select gene locus, says her colleague Dan Ryan, who works on CRISPR technologies. Customer orders vary — they might be for a 100-nucleotide sgRNA or a custom length and modification. He says the company's synthesis delivers high-quality gRNAs for extended-length gRNAs as well and that chemically synthesized RNAs perform better than the ones made in vitro from DNA templates. As CRISPR performance keeps improving and new applications emerge, says Bruhn, the widespread availability of CRISPR components "in new flavors" will bring these technologies to many labs. Agilent offers a gRNA design tool [online](#).

Integrated DNA Technologies (IDT) has a free-to-use CRISPR–Cas9 guide RNA [design checker](#) where labs can assess on- and off-target potential of a design for use on human, mouse, rat, zebrafish and *Caenorhabditis elegans* genomes. Earlier this year, the company launched a design tool for homology-directed repair (HDR) templates and Cas9 gRNAs. It is free to use with registration, says Justin Barr, who manages IDT's CRISPR and functional genomics products. He and his colleagues offer design help for guide RNAs that work with newer enzymes such as Cas13. For the design tools, the company built its own on-target gRNA prediction models, says Barr. And they designed and quality-controlled synthetic gRNAs. "We used in-house data for synthetic guides delivered into Cas9-expressing cells," he says. At the time, many tool developers were using lentiviral delivery of guides to obtain large datasets. An aspect people will want to keep in mind, he says, is the different performance between chemically synthesized guides delivered to a cell and those expressed in a cell from a DNA template ferried into the cell, such as by plasmid. This difference might be due to the temporal delay of transcription and translation before the RNA is available to complex with the Cas protein. This delay leads to larger and less controlled amounts of gRNA in the cell. With synthetic gRNA,

specific amounts of functional gRNA are delivered to the cell, often ‘pre-complexed’ with Cas9, and they can act quickly upon recognition of the target DNA sequence. He and his team have also observed that in vitro-generated gRNAs appear to cause more cellular toxicity, which may be due to carryover of contaminants from the in vitro preparation steps.

IDT uses electrospray ionization mass spec for gRNA quality control and to support high efficiency and avoid off-targets, says Barr. Another aspect to consider is contamination risk. Knowing how to synthesize nucleic acids does not automatically lead to reliable gRNAs. From “very savvy customers,” he says, “I’ve heard some pretty scary stories about the extent to which contamination can occur.” For knockout experiments, a lab might pick a sequence toward the beginning of a transcript and not worry much about off-targets, says Barr. With HDR, a lab navigates three variables in seeking a desirable on-target score, an edit close to the targeted window and low off-targets, he says. Labs can design and order chemically modified guides, such as those with phosphorothioated 2’ O-methyl bonds. Modifications stabilize the RNA, protect from exonuclease activity in a cell, help reduce an immune response, and lower the amount of material delivered to the cell, says Barr. Customers ask IDT how stable certain gRNAs and proteins in a ribonucleocomplex will be. The company does not sell such complexes but shares information from in-house research. “We have really done extensive stability studies, not just for the products that we’re selling,” he says.

Researchers can approach the company with custom requests, says Barr, such as to design specific aspects of an experiment for large projects in academia or industry, and for work in human or animal cells. Plants, too: “It’s growing, no pun intended, growing quickly,” he says. He also sees interest in pegRNAs for prime editing. It’s too early for IDT recommendations on pegRNAs, he says, but the company has been expanding its syntheses to make the long pegRNAs and handle challenges such as secondary structures. IDT’s pegRNAs are around 150 bases long, and some custom orders have been longer. He encourages labs to contact the IDT scientific applications team as they tinker with experimental design of more challenging projects.

Multiplexed gene-editing is a regular task at Qihan Biotech in Hangzhou, China. That’s when a number of genes in the same cells are edited. Luhan Yang cofounded Qihan with Harvard Medical School researcher George Church. She was a postdoctoral fellow in the Church lab and cofounded another company,



Prime editors have fundamentally changed how he and his team think about experiments, says Ben Kleinstiver, who took this photo. Among the Cas9 proteins his lab uses is SaCas9, a smaller Cas9 ortholog, seen here as a 3D-printed model. Credit: B. Kleinstiver, MGH

eGenesis, with him. At Qihan, Yang and her team genetically modify and differentiate stem cells to make donor cells for which a recipient might need less immune suppression. By editing out porcine endogenous retroviruses, they also study how pigs might more readily be a source of transplantable donor organs. In terms of online tools, to design gRNAs her team uses “whatever is handy,” she says, and follows up with internally developed algorithms. They check against the genome, filter out any gRNAs with more than four mismatches, and look at G+C content and homopolymers, she says. “Different algorithms have different priorities when they look at the guide RNA design parameters,” she says. In some cases, her team found efficiency prediction to be quite inaccurate, which is likely due to variable performance across cell types and genomic locations. When the Qihan team modifies many genes in parallel, the multiple DNA cuts carry an increased risk of on-target toxicity, rearrangements or translocations. On-target toxicity is largely due to the enzymatic reaction of multiple cuts to the genome, which stresses the cell, and “if we don’t pick the guide well, that will compound the problem.” Base editing, although limited in its possible types of genetic modifications, may be useful for multiplexed gene editing, says Yang. Single-stranded nicks are milder for a cell and present lower risks of translocations, she says. She has not yet tried prime editing, but sees its potential for “very precise editing.”

Base editing

Guide RNA design tools for base editing are less plentiful than for classic CRISPR–Cas9 systems. The lab of Sangsu Bae at Hanyang University in Korea has set up a web tool called **BE-Designer** for gRNA design in base

editing. Users can enter parameters and the tool will find target nucleotides and indicate aspects such as G+C content. **BE-Designer** shows potential off-target sites using **Cas-OFFinder**, also developed in the lab and available separately. For prime editing, the lab offers **PE-Designer** for pegRNA design for CRISPR prime editing. Kim says he finds these tools easy to use. With base editing, says Kleinstiver, the base of the intended edit must be in the edit window of the base editor construct, which is dictated by a PAM being available around 15 bases upstream from the target site. This is less of a limitation now that new Cas proteins target more PAM variants. It’s now generally possible to adequately position the edit window of the base editor, he says.

But both for base editing and for prime editing, says Kleinstiver, “you have a much narrower range of target sites that you can select from, since they have to be in close proximity to the intended edit.” When using these methods, he says, his team manually designs a series of target sites and then runs basic activity and specificity checks on them.

Prime editing

Prime editing, like base editing, nicks the DNA and writes changes in a genome. Base editors don’t intentionally make a double-strand break, but it happens a perceptible fraction of the time, says Kleinstiver, and the same is true for prime editing. At least, he says, this is based on the types of indels they leave, which are generally the result of double-stranded breaks, he says.

The pegRNA is a long gRNA that specifies the genomic target and encodes the desired edit. It uses nCas9, a Cas9 nickase fused to reverse transcriptase, to make genetic changes — insertions, deletions, point mutations and



"I am excited to hear that Nobel Prize is given to the field of my research," says Hyongbum Henry Kim (right), here with his lab. Credit: S. Park

combinations thereof. The developers call the method "a search and replace" genome editing technology⁸.

The pegRNA contains a guide sequence for target recognition, a tracrRNA scaffold sequence, a primer binding site (PBS) and a reverse transcription template (RTT) that sets the reverse transcriptase in motion. As the method's developers note, they devised a way to nick DNA at the target site to expose a 3'-hydroxyl group, which can be used to prime the reverse transcription of an edit-encoding extension to be written directly into the target site. One type of prime editor, PE1, has a pegRNA and the RNA-programmable nCas9 fusion. PE2 has an engineered reverse transcriptase that yields better editing efficiency and PE3 involves an accessory nicking gRNA that nicks the non-edited strand to induce its replacement, which further raises editing efficiency, says Kleinstiver. As Kim explains, nicking the non-edited strand in PE3 is similar to nicking the non-edited strand in base editors. "Prime editors have fundamentally changed how we think about experiments in the lab, including for making heterozygous cell lines," says Kleinstiver. This was previously not a trivial task, given that generating these lines by HDR often resulted in an unwanted insertion or deletion mutation on one allele. Base-editing users might also find unintended bystander edits, says Kleinstiver. In prime editing, he says, PBS or RTT length or PE3 gRNA composition can widely affect editing efficiency. It makes prime editing somewhat complex that "there are lot of moving parts," he says, and that can make experimental design "a bit intimidating." A lab has to select the initial target site spacer to be used to guide the prime editor construct to the

target. That is similar to the spacer with any other sgRNA for CRISPR editing. There can be several target site options, but pegRNAs will have different activity at different sites or positions of the edit at the target site. That affects prime-editing efficiency. Another consideration for researchers, he says, is the accessory nicking gRNA, such as PE3 or PE3b gRNA, which directs the nCas9 enzyme in the prime editor construct toward nicking the non-edited strand. This is done to 'trick' DNA repair enzymes into incorporating the change embedded in the pegRNA.

Kim sees prime editing as a promising genome-editing method and finds it clever that the gRNA encodes the information for the intended edit. A small genetic edit calls for only two components: the pegRNA and, for example, PE2. Kleinstiver points to pegRNA design tools such as [PrimeDesign](#) and [pegFinder](#). Kim's group has developed pegRNA guidance that "will certainly help," says Doench. From a principles standpoint, he says 'rules' can help bias researchers toward success. With a specific envisioned edit, labs will want to try a few constructs in parallel and use an assay to find the best among them. This is usually PCR followed by sequencing. In this way, prime editing is not so unlike "classic" CRISPR editing and HDR. It's advisable, he says, to try a few guides and repair templates at the same time. Kim and his team have explored the factors that shape prime-editing efficiencies and developed three computational models to predict how efficient pegRNAs might be⁹. Their tools to do so are DeepPE, PE_type and PE_position, available [online](#). Kim says these models predict pegRNA activity but they do not predict off-target effects. More work lies ahead, but he expects prime

editing to have limited off-targets. Despite the length of pegRNAs, which could create secondary structures, he and his lab find the method works well. Some of the pegRNA recommendations from the Kim lab include using a 13-nucleotide primer binding site and 12-nucleotide reverse transcriptase template, looking for high G+C count in the PBS region when possible, using a G as the last templated nucleotide when the reverse transcriptase template length is less than 12 nucleotides, and assuring that the GG sequence in the PAM of the target sequence is modified so that Cas9 nickase cannot bind again to the edited sequence.

Doench has not yet tried these models to design pegRNAs, "but I'd certainly think that users will be better off with these predictions than with random guessing about how to tune all the different knobs." Just like other tools, these are "not going to be perfect, but should still bias towards success." Over time, more biasing toward success will happen and rules will be established, he says. The challenge, in his view, is that prime editing is "still not yet a very efficient process in some-to-many cell types." This means a scientist might need to do single-cell cloning to get a pure population, "and that's a process that is, at best, annoying," says Doench. The same is true for HDR-mediated editing, "so really, I'm blaming the cells in this case, not CRISPR technology."

"Prime editing is super cool," says Mali. "Really the biggest issue I see is delivery given the huge payload." He has not yet tried the tools from the Kim lab but says that the design rules seem reasonable and that they build on the Kim lab's work on gRNA design and optimization. gRNA design tools are valuable both for enabling one-off studies and for designing libraries for large-scale screens, he says. As more experimental and training data are accrued, he says, these resources "will only get better, and really further democratize the CRISPR toolsets." □

Vivien Marx[✉]
Nature Methods.
[✉]e-mail: v.marx@us.nature.com

Published online: 5 November 2020
<https://doi.org/10.1038/s41592-020-01003-4>

References

- Chen, B. et al. *Cell* **155**, 1479–1491 (2013).
- Doench, J. G. et al. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
- Kiani, S. et al. *Nat. Methods* <https://doi.org/10.1038/nmeth.3580> (2015).
- Michlits, G. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-020-0850-8> (2020).
- Hanna, R. E. & Doench, J. G. *Nat. Biotechnol.* **38**, 813–823 (2020).
- Jinek, M. et al. *Science* **337**, 816–821 (2012).
- Kleinstiver, B. et al. *Nat. Biotechnol.* **37**, 276–282 (2019).
- Anzalone, A. V. et al. *Nature* **576**, 149–157 (2019).
- Kim, H. K. et al. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0677-y> (2020).