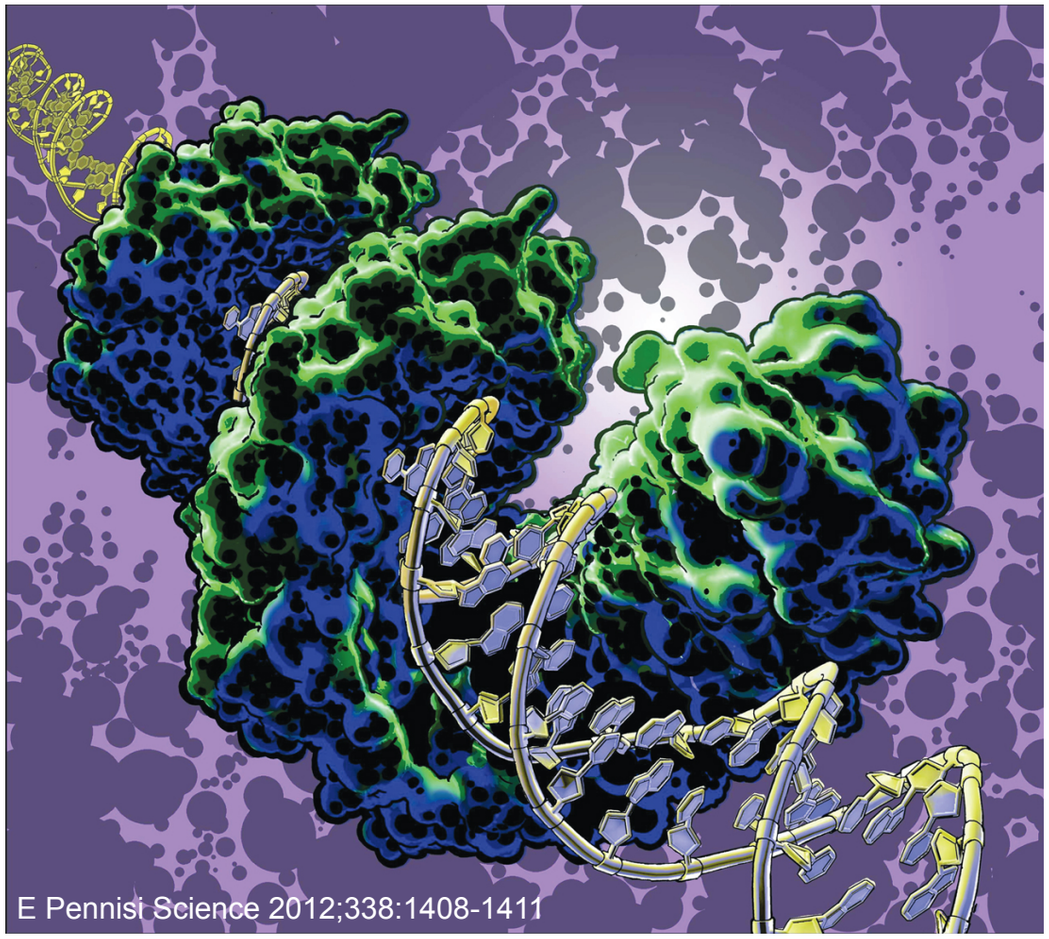
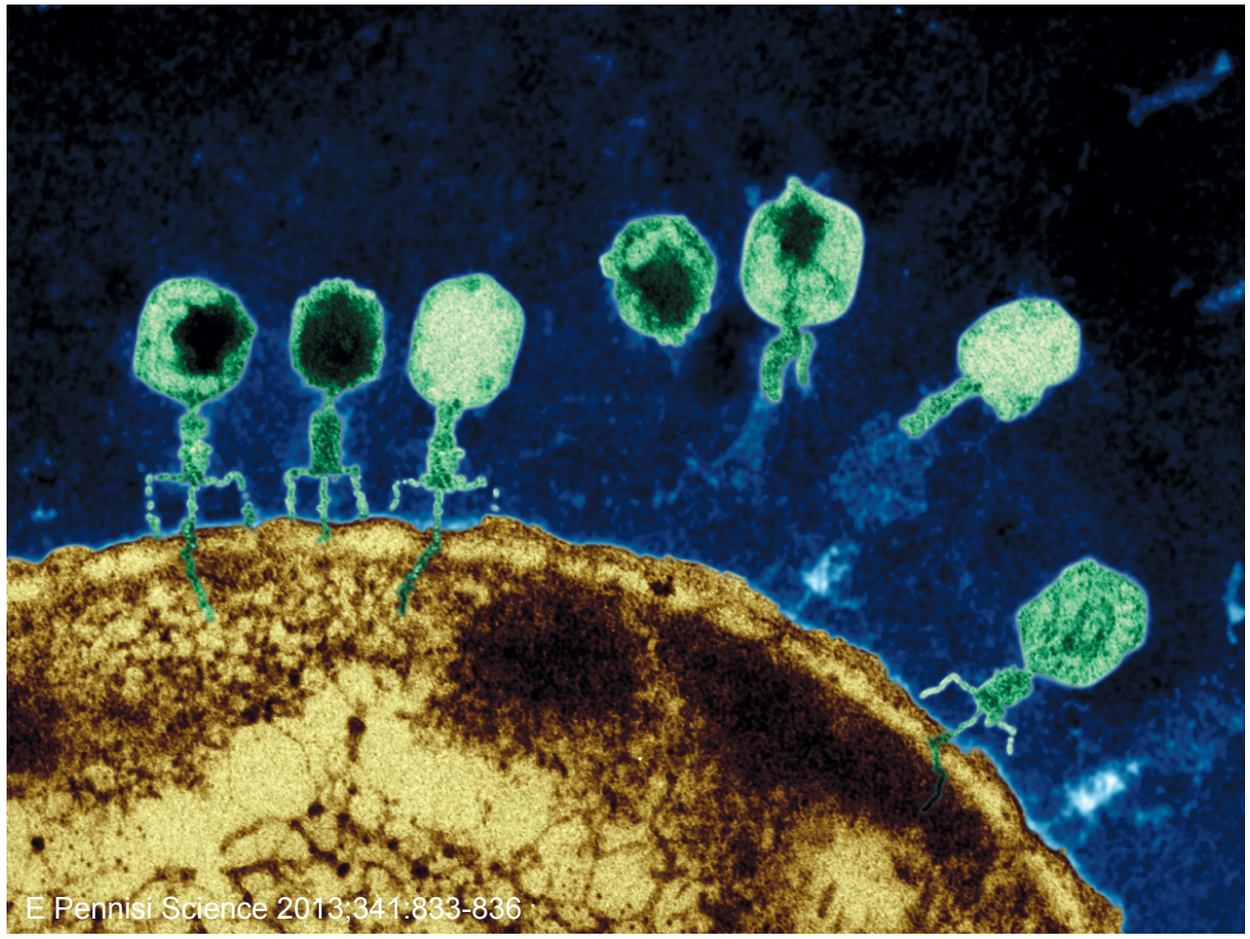
Module 9: Genome Editing and Regulation

CSE 590: Molecular programming and neural computation

Guest Lecture: Nick Bogard

Genome Engineering



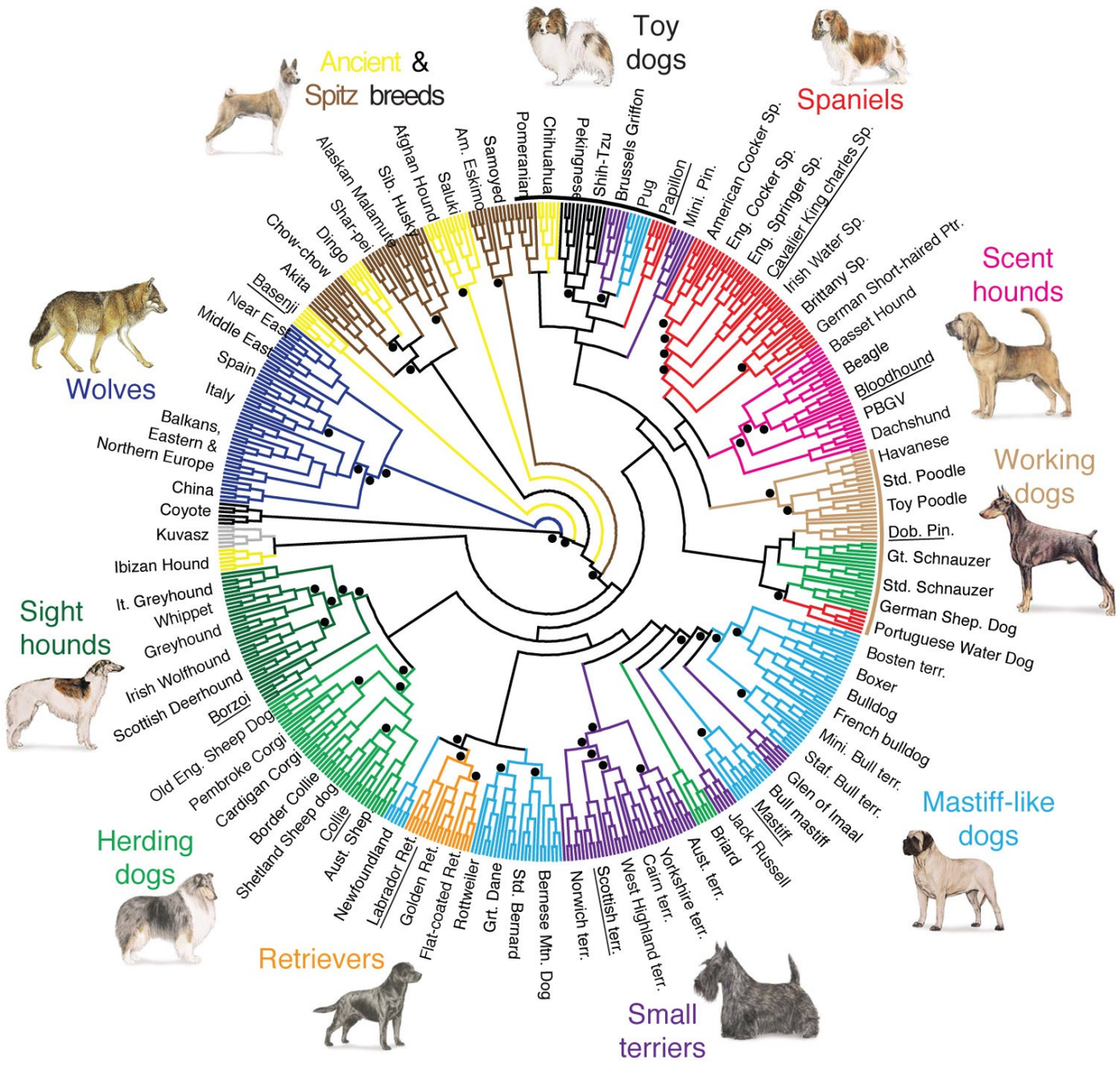


"Talking about the future is better than letting it sneak up on us. We need to do more of this or we will be left with very limited vocabulary in the space between positive and negative hype."

- George Church, speaking on the potential of CRISPR technology

Genome “Engineering”

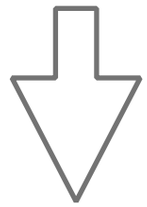
http://www.synthetic-bestiary.com/



http://newsroom.ucla.edu/

Genome Engineering

Cut the target

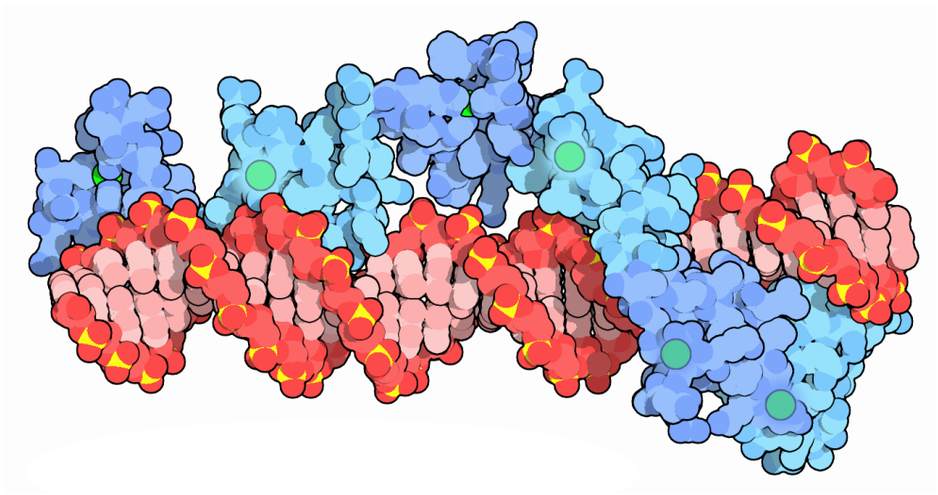


Add modification

New Genomic Sequence

Genome Editing

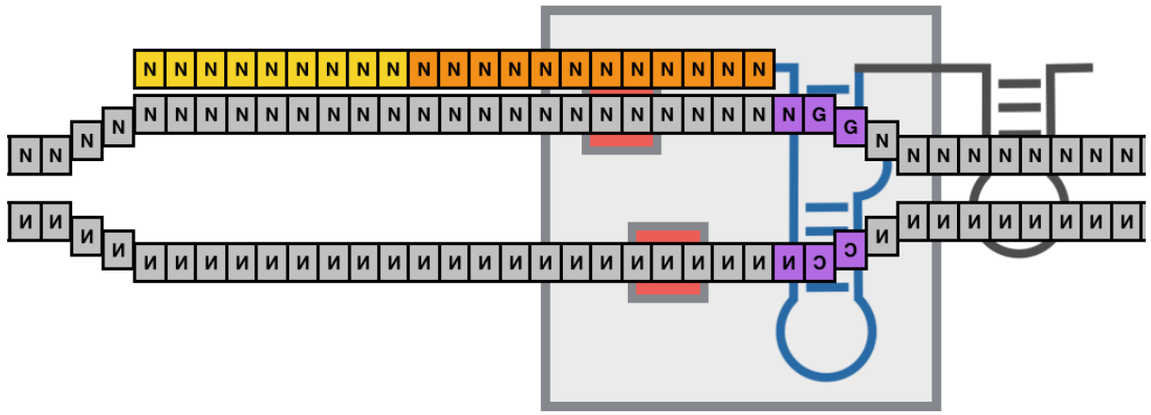
Protein-based



1. Protein and DNA interface is very complex

2. Construction is relatively expensive and complicated

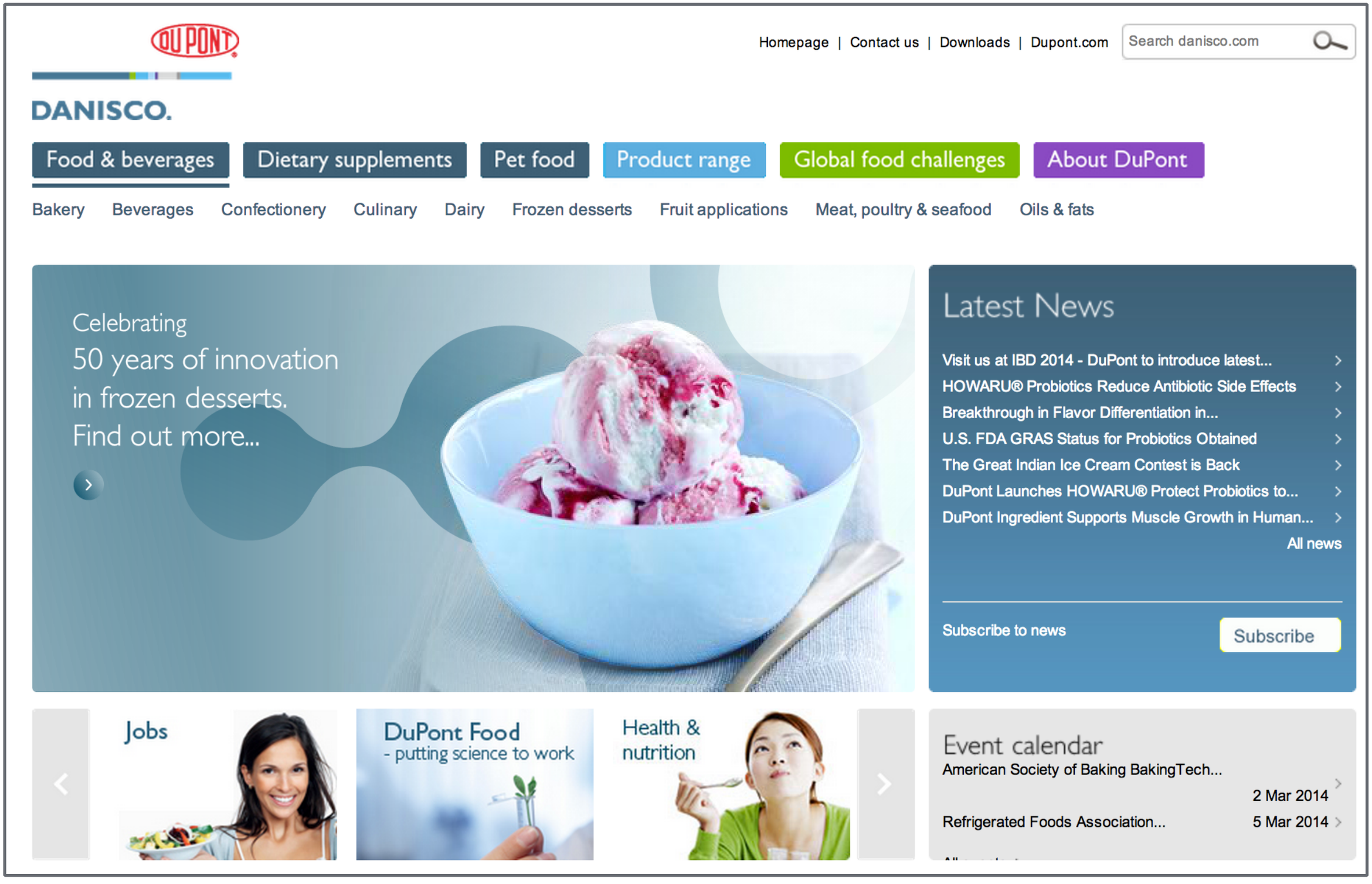
CRISPR



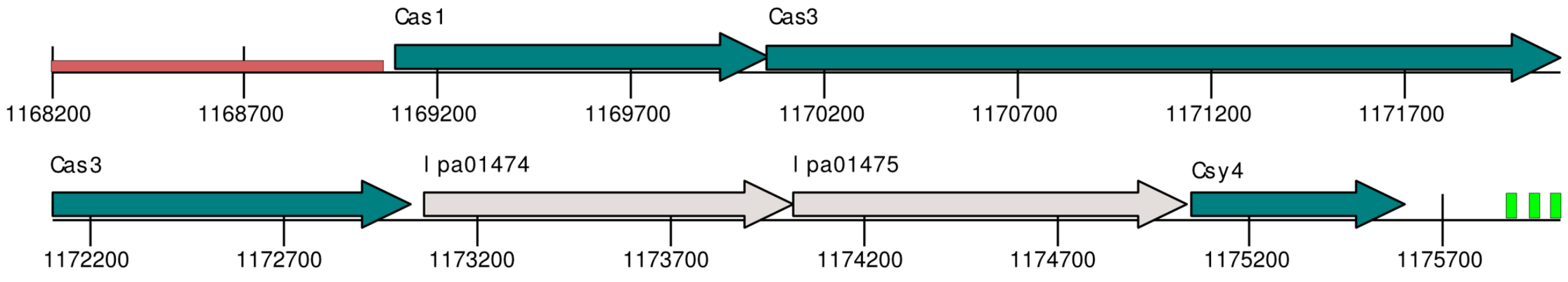
1. Watson-Crick base pairing is straight-forward and predictable

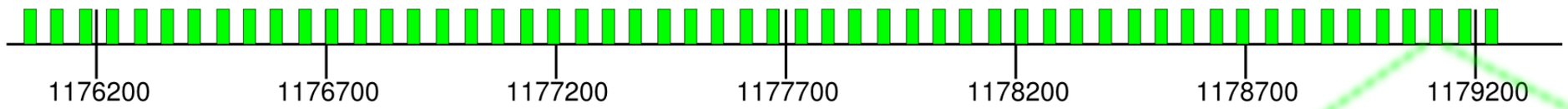
2. Construction is fast, easy, and cheap

Bacterial Immunity



Bacterial Immunity





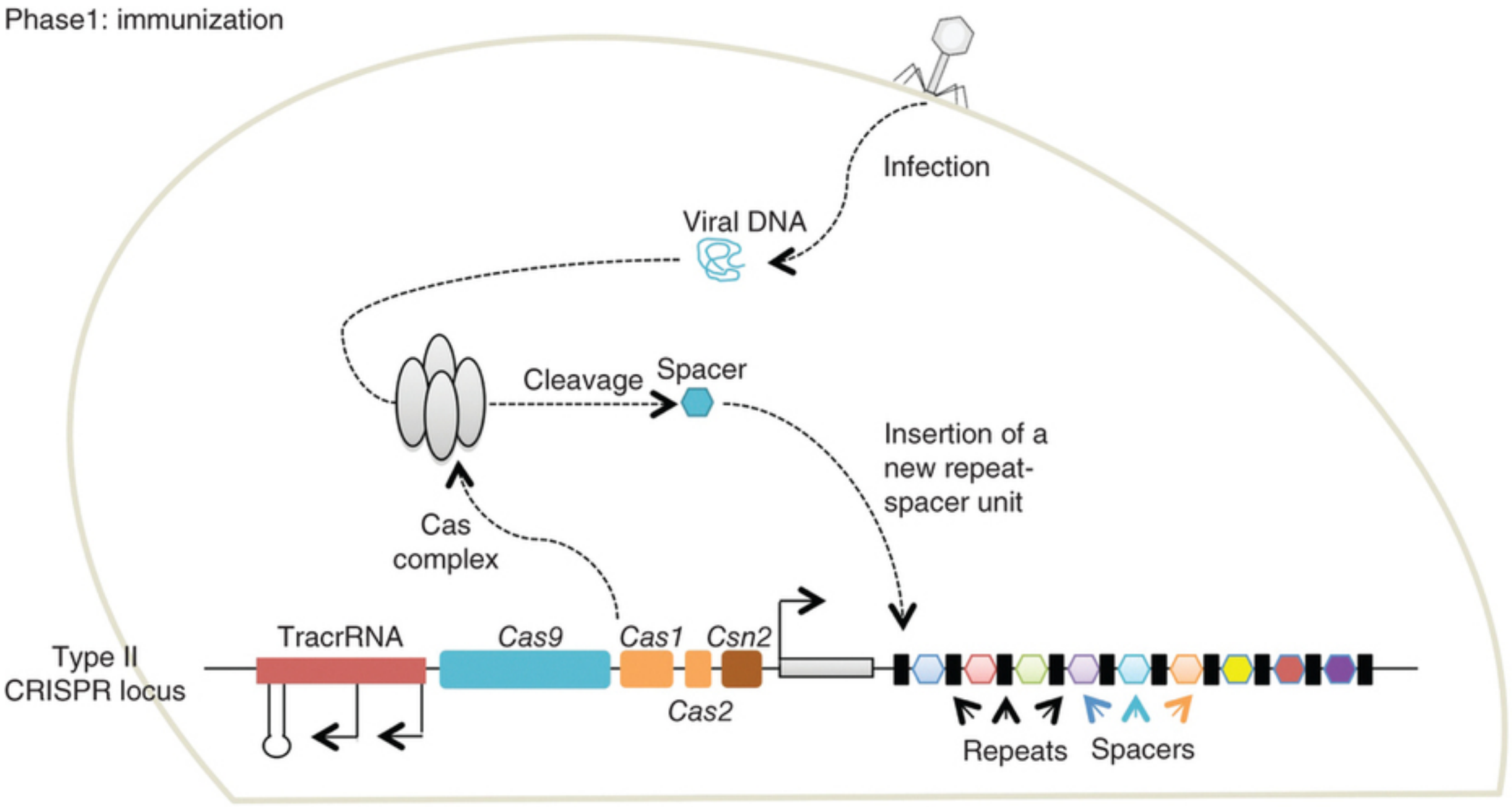
CRISPR Associated proteins
Clustered Regularly Interspaced Short Palindromic Repeats
Hypothetical proteins
AT rich region

gttaactgccgcacaggcagcttagaa

Repeat sequence. In red is indicated
the unique difference with the Lens CRISPRs

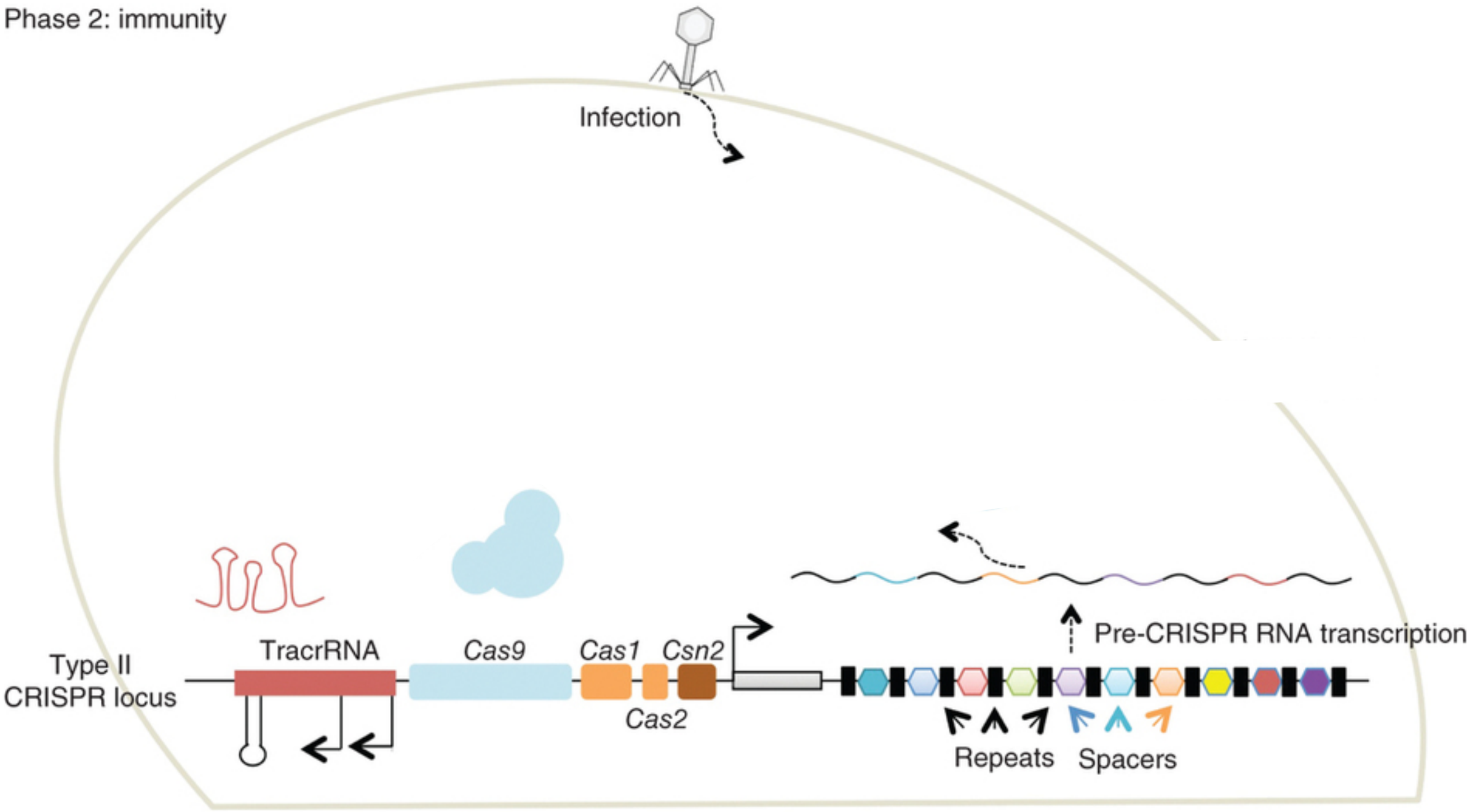
D'Auria et al. *BMC Genomics* 2010

Bacterial Immunity



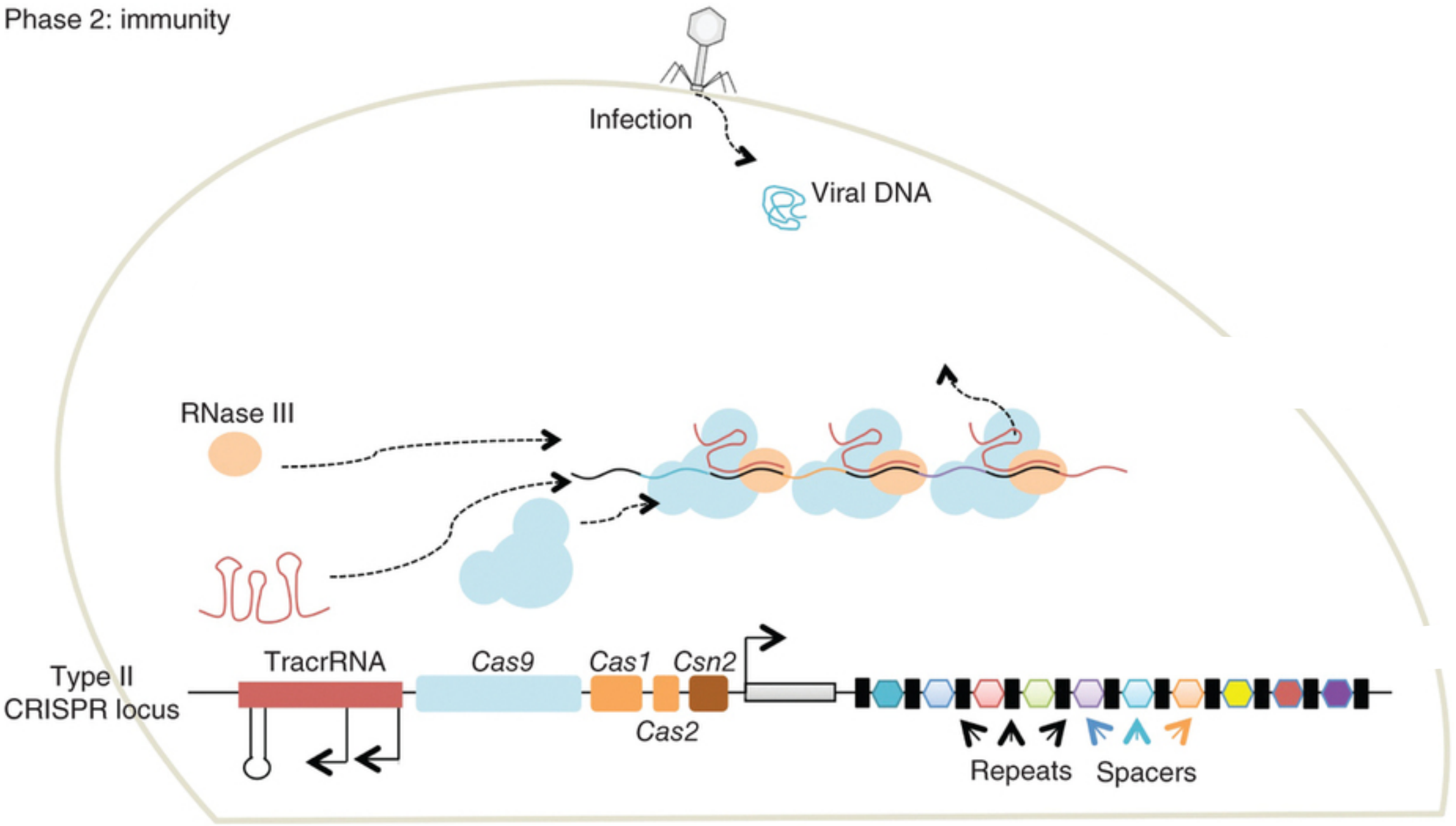
Mali et al. *Nature Methods* 2013

Bacterial Immunity



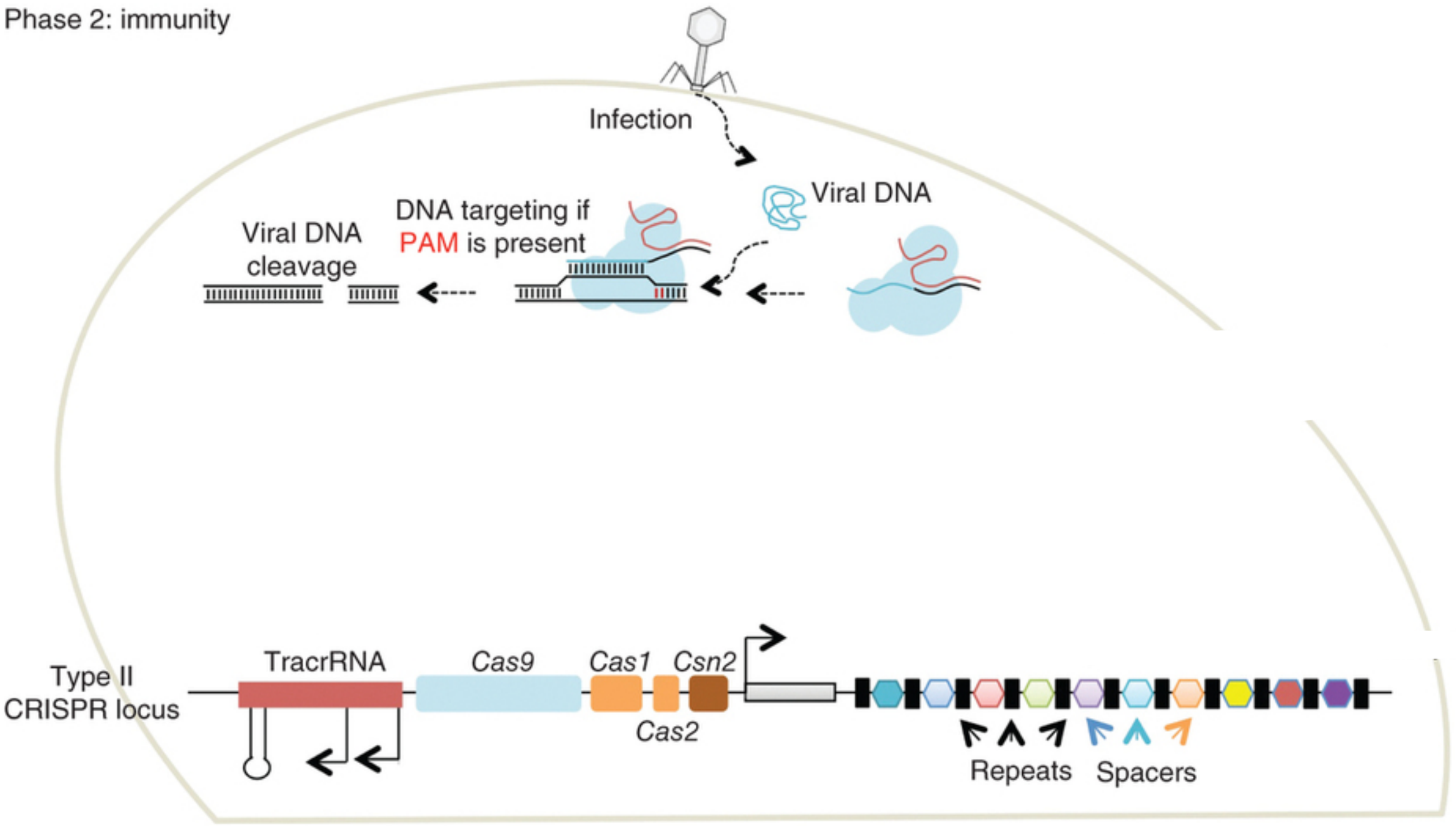
Mali et al. *Nature Methods* 2013

Bacterial Immunity



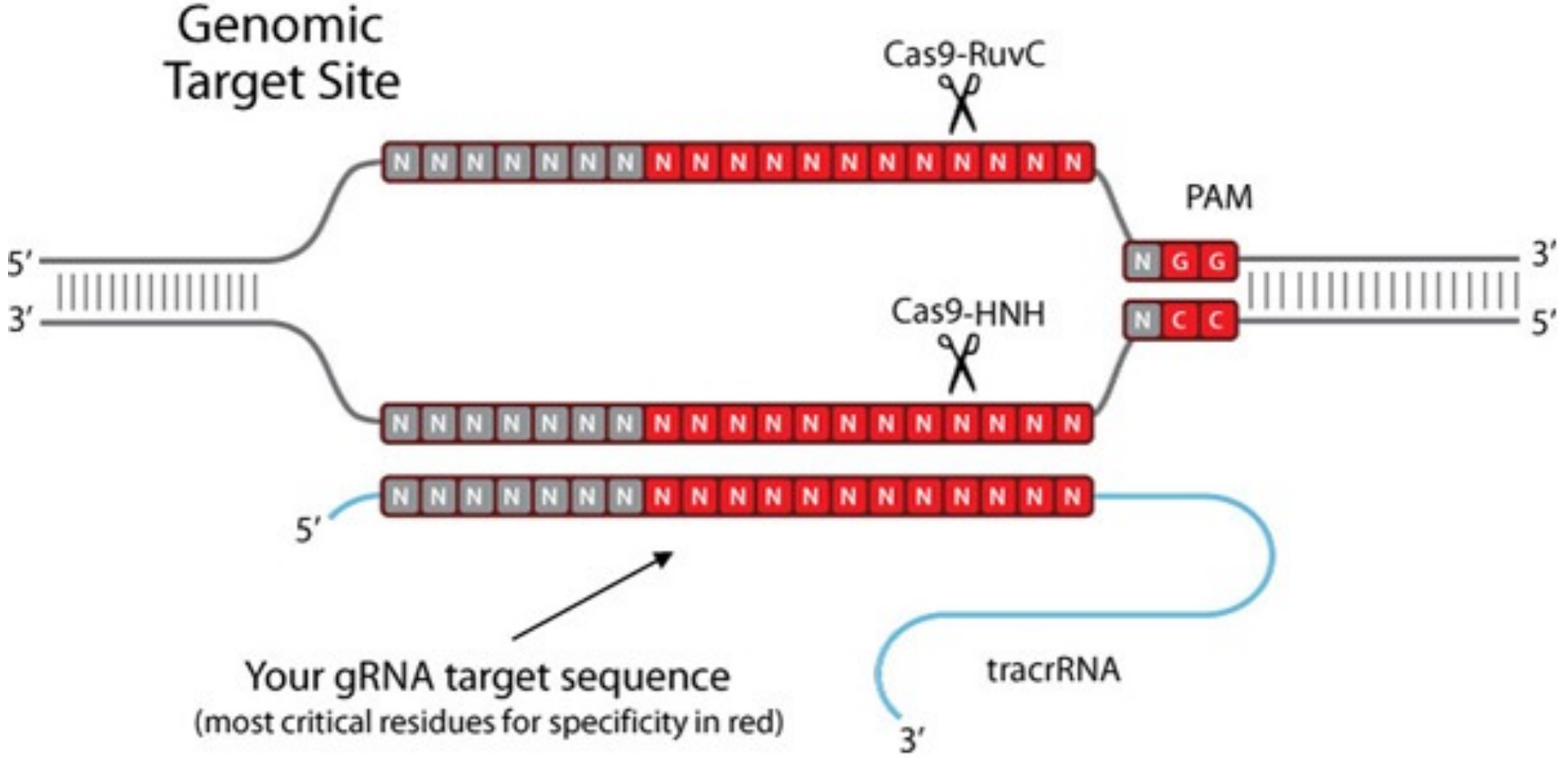
Mali et al. *Nature Methods* 2013

Bacterial Immunity

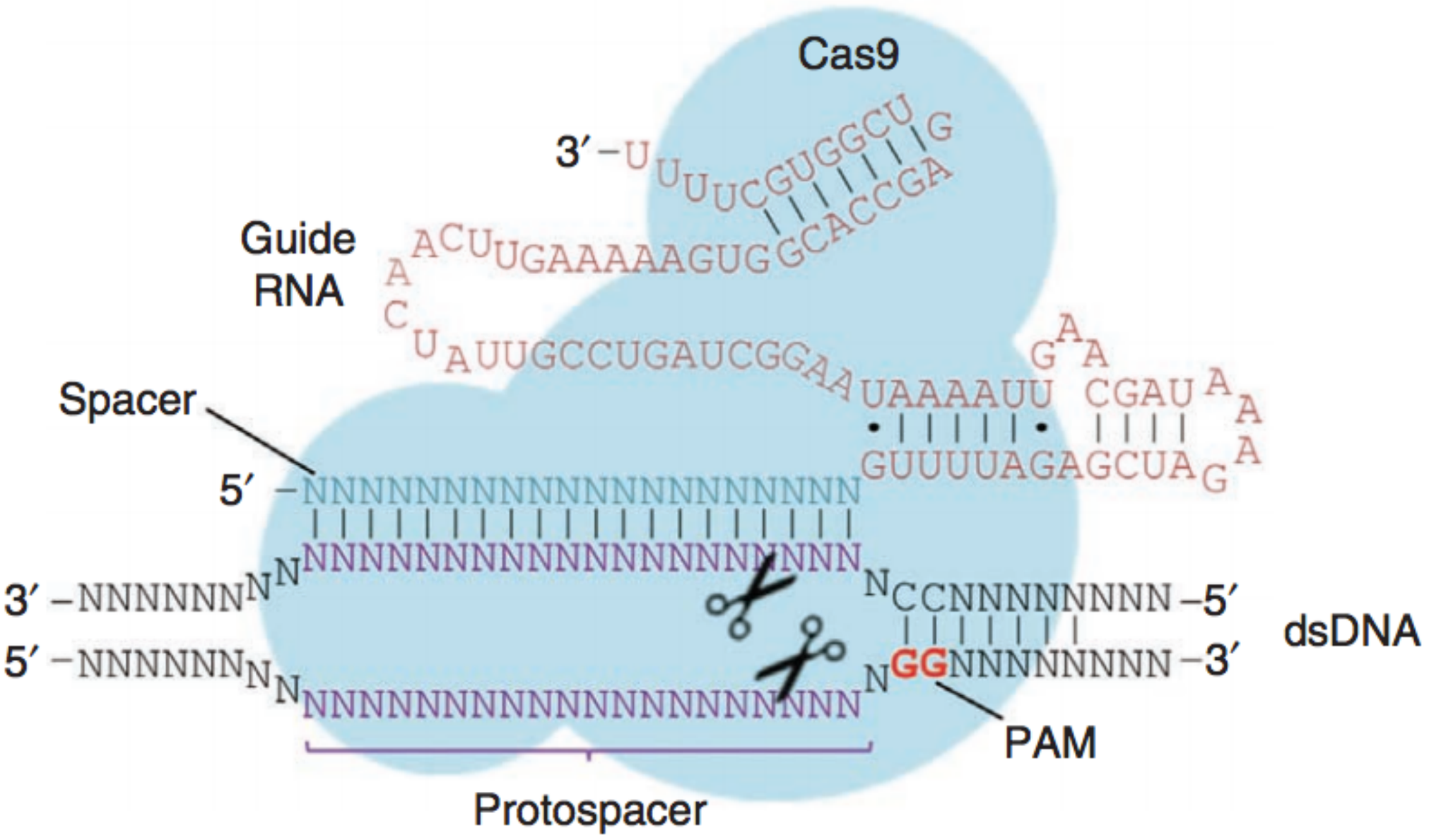


Mali et al. *Nature Methods* 2013

CRISPR Targeting



CRISPR/Cas/DNA Complex



Mali et al. *Nature Methods* 2013

Genome Engineering

RESEARCH ARTICLE

A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity

Martin Jinek,[1,2]* Krzysztof Chylinski,[3,4]* Ines Fonfara,[4] Michael Hauer,[2]† Jennifer A. Doudna,[1,2,5,6]‡ Emmanuelle Charpentier[4]‡

Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems provide bacteria and archaea with adaptive immunity against viruses and plasmids by using CRISPR RNAs (crRNAs) to guide the silencing of invading nucleic acids. We show here that in a subset of these systems, the mature crRNA that is base-paired to trans-activating crRNA (tracrRNA) forms a two-RNA structure that directs the CRISPR-associated protein Cas9 to introduce double-stranded (ds) breaks in target DNA. At sites complementary to the crRNA-guide sequence, the Cas9 HNH nuclease domain cleaves the complementary strand, whereas the Cas9 RuvC-like domain cleaves the noncomplementary strand. The dual-tracrRNA:crRNA, when engineered as a single RNA chimera, also directs sequence-specific Cas9 dsDNA cleavage. Our study reveals a family of endonucleases that use dual-RNAs for site-specific DNA cleavage and highlights the potential to exploit the system for RNA-programmable genome editing.

Bacteria and archaea have evolved RNA-mediated adaptive defense systems called clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) that protect organisms from invading viruses and plasmids (*1*–*3*). These defense systems rely on small RNAs for sequence-specific detection and silencing of foreign nucleic acids. CRISPR/Cas systems are composed of *cas* genes organized in operon(s) and CRISPR array(s) consisting of genome-targeting sequences (called spacers) interspersed with identical repeats (*1*–*3*). CRISPR/Cas-mediated immunity occurs in three steps. In the adaptive phase, bacteria and archaea harboring one or more CRISPR loci respond to viral or plasmid challenge by integrating short fragments of foreign sequence (protospacers) into the host chromosome at the proximal end of the CRISPR array (*1*–*3*). In the expression and interference phases, transcription of the repeat-spacer element into precursor CRISPR RNA (pre-crRNA) molecules followed by enzymatic cleavage yields the short crRNAs that can pair with complementary protospacer sequences of invading viral or plasmid targets (*4*–*11*). Target recognition by crRNAs directs the silencing of the foreign sequences by means of Cas proteins that function in complex with the crRNAs (*10*, *12*–*20*).

There are three types of CRISPR/Cas systems (*21*–*23*). The type I and III systems share some overarching features: specialized Cas endonucleases process the pre-crRNAs, and once mature, each crRNA assembles into a large multi-Cas protein complex capable of recognizing and cleaving nucleic acids complementary to the crRNA. In contrast, type II systems process pre-crRNAs by a different mechanism in which a trans-activating crRNA (tracrRNA) complementary to the repeat sequences in pre-crRNA triggers processing by the double-stranded (ds) RNA-specific ribonuclease RNase III in the presence of the Cas9 (formerly Csn1) protein (fig. S1) (*4*, *24*). Cas9 is thought to be the sole protein responsible for crRNA-guided silencing of foreign DNA (*25*–*27*).

We show here that in type II systems, Cas9 proteins constitute a family of enzymes that require a base-paired structure formed between the activating tracrRNA and the targeting crRNA to cleave target dsDNA. Site-specific cleavage occurs at locations determined by both base-pairing complementarity between the crRNA and the target protospacer DNA and a short motif [referred to as the protospacer adjacent motif (PAM)] juxtaposed to the complementary region in the target DNA. Our study further demonstrates that the Cas9 endonuclease family can be programmed with single RNA molecules to cleave specific DNA sites, thereby raising the exciting possibility of developing a simple and versatile RNA-directed system to generate dsDNA breaks for genome targeting and editing.

Cas9 is a DNA endonuclease guided by two RNAs. Cas9, the hallmark protein of type II systems, has been hypothesized to be involved in both crRNA maturation and crRNA-guided DNA interference (fig. S1) (*4*, *25*–*27*). Cas9 is involved in crRNA maturation (*4*), but its direct participation in target DNA destruction has not been investigated. To test whether and how Cas9 might be capable of target DNA cleavage, we used an overexpression system to purify Cas9 protein derived from the pathogen *Streptococcus pyogenes* (fig. S2, see supplementary materials and methods) and tested its ability to cleave a plasmid DNA or an oligonucleotide duplex bearing a protospacer sequence complementary to a mature crRNA, and a bona fide PAM. We found that mature crRNA alone was incapable of directing Cas9-catalyzed plasmid DNA cleavage (Fig. 1A and fig. S3A). However, addition of tracrRNA, which can pair with the repeat sequence of crRNA and is essential to crRNA maturation in this system, triggered Cas9 to cleave plasmid DNA (Fig. 1A and fig. S3A). The cleavage reaction required both magnesium and the presence of a crRNA sequence complementary to the DNA; a crRNA capable of tracrRNA base pairing but containing a noncognate target DNA-binding sequence did not support Cas9-catalyzed plasmid cleavage (Fig. 1A; fig. S3A, compare crRNA-sp2 to crRNA-sp1; and fig. S4A). We obtained similar results with a short linear dsDNA substrate (Fig. 1B and fig. S3, B and C). Thus, the trans-activating tracrRNA is a small noncoding RNA with two critical functions: triggering pre-crRNA processing by the enzyme RNase III (*4*) and subsequently activating crRNA-guided DNA cleavage by Cas9.

Cleavage of both plasmid and short linear dsDNA by tracrRNA:crRNA-guided Cas9 is site-specific (Fig. 1, C to E, and fig. S5, A and B). Plasmid DNA cleavage produced blunt ends at a position three base pairs upstream of the PAM sequence (Fig. 1, C and E, and fig. S5, A and C) (*26*). Similarly, within short dsDNA duplexes, the DNA strand that is complementary to the target-binding sequence in the crRNA (the complementary strand) is cleaved at a site three base pairs upstream of the PAM (Fig. 1, D and E, and fig. S5, B and C). The noncomplementary DNA strand is cleaved at one or more sites within three to eight base pairs upstream of the PAM. Further investigation revealed that the noncomplementary strand is first cleaved endonucleolytically and subsequently trimmed by a 3′-5′ exonuclease activity (fig. S4B). The cleavage rates by Cas9 under single-turnover conditions ranged from 0.3 to 1 min^{-1}, comparable to those of restriction endonucleases (fig. S6A), whereas incubation of wild-type (WT) Cas9-tracrRNA:crRNA complex with a fivefold molar excess of substrate DNA provided evidence that the dual-RNA–guided Cas9 is a multiple-turnover enzyme (fig. S6B). In

[1]Howard Hughes Medical Institute (HHMI), University of California, Berkeley, CA 94720, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [3]Max F. Perutz Laboratories (MFPL), University of Vienna, A-1030 Vienna, Austria. [4]The Laboratory for Molecular Infection Medicine Sweden, Umeå Centre for Microbial Research, Department of Molecular Biology, Umeå University, S-90187 Umeå, Sweden. [5]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [6]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.
†Present address: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.
‡To whom correspondence should be addressed. E-mail: doudna@berkeley.edu (J.A.D.); emmanuelle.charpentier@mims.umu.se (E.C.)

4. A. J. Wood *et al.*, *Science* **333**, 307 (2011).
5. M. Christian *et al.*, *Genetics* **186**, 757 (2010).
6. F. Zhang *et al.*, *Nat. Biotechnol.* **29**, 149 (2011).
7. J. C. Miller *et al.*, *Nat. Biotechnol.* **29**, 143 (2011).
8. D. Reyon *et al.*, *Nat. Biotechnol.* **30**, 460 (2012).
9. J. Boch *et al.*, *Science* **326**, 1509 (2009).
10. M. J. Moscou, A. J. Bogdanove, *Science* **326**, 1501 (2009).
11. B. L. Stoddard, *Q. Rev. Biophys.* **38**, 49 (2005).
12. M. Jinek *et al.*, *Science* **337**, 816 (2012).
13. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579 (2012).
14. J. E. Garneau *et al.*, *Nature* **468**, 67 (2010).
15. H. Deveau, J. E. Garneau, S. Moineau, *Annu. Rev. Microbiol.* **64**, 475 (2010).
16. P. Horvath, R. Barrangou, *Science* **327**, 167 (2010).
17. K. S. Makarova *et al.*, *Nat. Rev. Microbiol.* **9**, 467 (2011).
18. D. Bhaya, M. Davison, R. Barrangou, *Annu. Rev. Genet.* **45**, 273 (2011).
19. E. Deltcheva *et al.*, *Nature* **471**, 602 (2011).
20. R. Sapranauskas *et al.*, *Nucleic Acids Res.* **39**, 9275 (2011).
21. A. H. Magadán, M. E. Dupuis, M. Villion, S. Moineau, *PLoS ONE* **7**, e40913 (2012).
22. H. Deveau *et al.*, *J. Bacteriol.* **190**, 1390 (2008).
23. F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, *Microbiology* **155**, 733 (2009).
24. M. Jinek, J. A. Doudna, *Nature* **457**, 405 (2009).
25. C. D. Malone, G. J. Hannon, *Cell* **136**, 656 (2009).
26. G. Meister, T. Tuschl, *Nature* **431**, 343 (2004).
27. M. T. Certo *et al.*, *Nat. Methods* **8**, 671 (2011).
28. P. Mali *et al.*, *Science* **339**, 823 (2013).
29. P. A. Carr, G. M. Church, *Nat. Biotechnol.* **27**, 1151 (2009).

Acknowledgments: We thank the entire Zhang lab for their support and advice; P. A. Sharp for generous help with Northern blot analysis; C. Jennings, R. Desimone, and M. Kowalczyk for helpful comments; and X. Ye for help with confocal imaging. L.C. and X.W. are Howard Hughes Medical Institute International Student Research Fellows. D.C. is supported by the Medical Scientist Training Program. P.D.H. is a James Mills Pierce Fellow. X.W. is supported by NIH grants R01-GM34277 and R01-CA133404 to P. A. Sharp, X.W.'s thesis adviser. L.A.M. is supported by Searle Scholars, R. Allen, an Irma T. Hirschl Award, and a NIH Director's New Innovator Award (DP2AI104556). F.Z. is supported by a NIH Director's Pioneer Award (DP1MH100706); the Keck, McKnight, Gates, Damon Runyon, Searle Scholars, Klingenstein, and Simons foundations; R. Metcalfe; M. Boylan; and J. Pauley. The authors have no conflicting financial interests. A patent application has been filed relating to this work, and the authors plan on making the reagents widely available to the academic community through Addgene and to provide software tools via the Zhang lab Web site (www.genome-engineering.org).

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1231143/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 and S2
References (*30*–*32*)

5 October 2012; accepted 12 December 2012
Published online 3 January 2013;
10.1126/science.1231143

RNA-Guided Human Genome Engineering via Cas9

Prashant Mali,[1]* Luhan Yang,[1,3]* Kevin M. Esvelt,[2] John Aach,[1] Marc Guell,[1] James E. DiCarlo,[4] Julie E. Norville,[1] George M. Church[1,2]†

Bacteria and archaea have evolved adaptive immune defenses, termed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems, that use short RNA to direct degradation of foreign nucleic acids. Here, we engineer the type II bacterial CRISPR system to function with custom guide RNA (gRNA) in human cells. For the endogenous AAVS1 locus, we obtained targeting rates of 10 to 25% in 293T cells, 13 to 8% in K562 cells, and 2 to 4% in induced pluripotent stem cells. We show that this process relies on CRISPR components; is sequence-specific; and, upon simultaneous introduction of multiple gRNAs, can effect multiplex editing of target loci. We also compute a genome-wide resource of ~190 K unique gRNAs targeting ~40.5% of human exons. Our results establish an RNA-guided editing tool for facile, robust, and multiplexable human genome engineering.

Bacterial and archaeal clustered regularly interspaced short palindromic repeats (CRISPR) systems rely on CRISPR RNAs (crRNAs) in complex with CRISPR-associated (Cas) proteins to direct degradation of complementary sequences present within invading viral and plasmid DNA (*1*–*3*). A recent in vitro reconstitution of the *Streptococcus pyogenes* type II CRISPR system demonstrated that crRNA fused to a normally trans-encoded tracrRNA is sufficient to direct Cas9 protein to sequence-specifically cleave target DNA sequences matching the crRNA (*4*). The fully defined nature of this two-component system suggested that it might function in the cells of eukaryotic organisms such as yeast, plants, and even mammals. By cleaving genomic sequences targeted by RNA sequences (*4*–*6*), such a system could greatly enhance the ease of genome engineering.

Here, we engineer the protein and RNA components of this bacterial type II CRISPR system in human cells. We began by synthesizing a human codon–optimized version of the Cas9 protein bearing a C-terminal SV40 nuclear localization signal and cloning it into a mammalian expression system (Fig. 1A and fig. S1A). To direct Cas9 to cleave sequences of interest, we expressed crRNA-tracrRNA fusion transcripts, hereafter referred to as guide RNAs (gRNAs), from the human U6 polymerase III promoter. Directly transcribing gRNAs allowed us to avoid reconstituting the RNA-processing machinery used by bacterial CRISPR systems (Fig. 1A and fig. S1B) (*4*, *7*–*9*). Constrained only by U6 transcription initiating with G and the requirement for the PAM (protospacer-adjacent motif) sequence -NGG following the 20–base pair (bp) crRNA target, our highly versatile approach can, in principle, target any genomic site of the form $GN_{20}GG$ (fig. S1C; see supplementary text S1 for a detailed discussion).

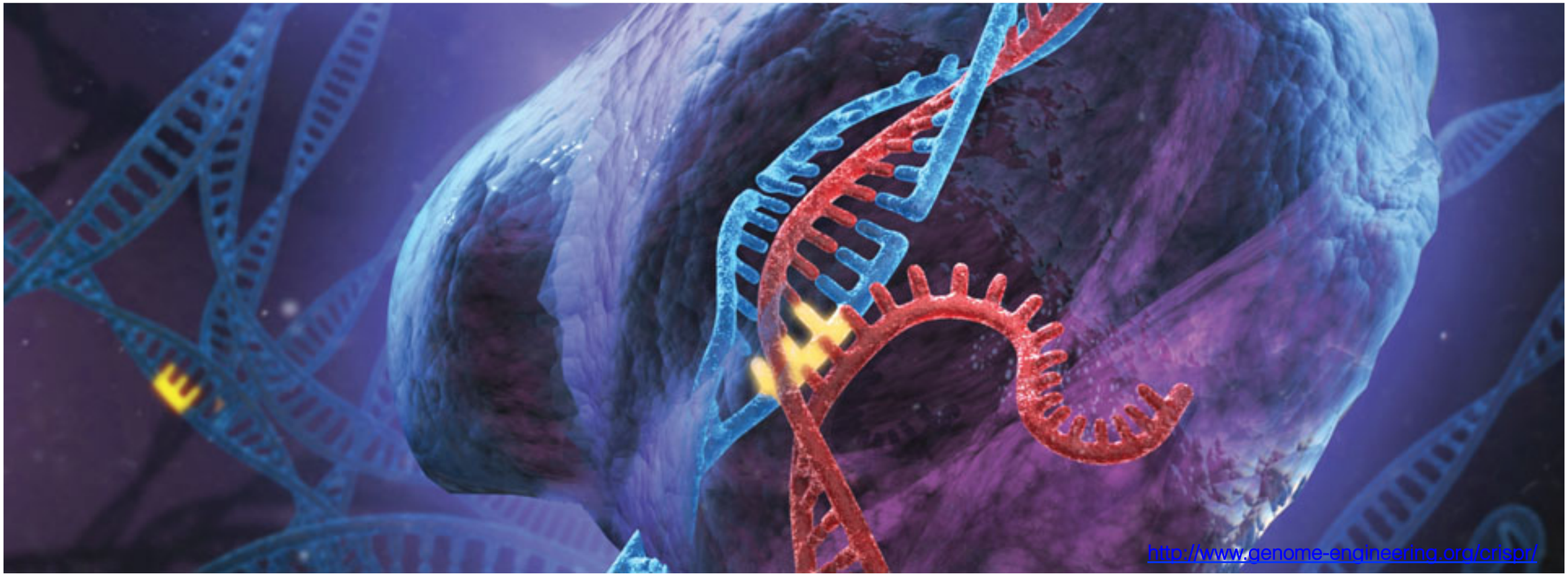
To test the functionality of our implementation for genome engineering, we developed a green fluorescent protein (GFP) reporter assay (Fig. 1B) in human embryonic kidney HEK 293T cells similar to one previously described (*10*). Specifically, we established a stable cell line bearing a genomically integrated GFP coding sequence disrupted by the insertion of a stop codon and a 68-bp genomic fragment from the AAVS1 locus that renders the expressed protein fragment nonfluorescent. Homologous recombination (HR) using an appropriate repair donor can restore the normal GFP sequence, which enabled us to quantify the resulting GFP^+ cells by flow-activated cell sorting (FACS).

To test the efficiency of our system at stimulating HR, we constructed two gRNAs, T1 and T2, that target the intervening AAVS1 fragment (Fig. 1B) and compared their activity to that of a previously described TAL effector nuclease heterodimer (TALEN) targeting the same region (*11*). We observed successful HR events using all three targeting reagents, with gene correction rates using the T1 and T2 gRNAs approaching 3% and 8%, respectively (Fig. 1C). This RNA-mediated editing process was notably rapid, with the first detectable GFP^+ cells appearing ~20 hours post transfection compared with ~40 hours for the AAVS1 TALENs. We observed HR only upon simultaneous introduction of the repair donor, Cas9 protein, and gRNA, which confirmed that all components are required for genome editing (fig. S2). Although we noted no apparent toxicity associated with Cas9/gRNA expression, work with zinc finger nucleases (ZFNs) and TALENs has shown that nicking only one strand further reduces toxicity. Accordingly, we also tested a Cas9D10A mutant that is known to function as a nickase in vitro, which yielded similar HR but lower nonhomologous end joining (NHEJ) rates (fig. S3) (*4*, *5*). Consistent with (*4*), in which a related Cas9 protein is shown to cut both strands

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA. [3]Biological and Biomedical Sciences Program, Harvard Medical School, Boston, MA 02115, USA. [4]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.

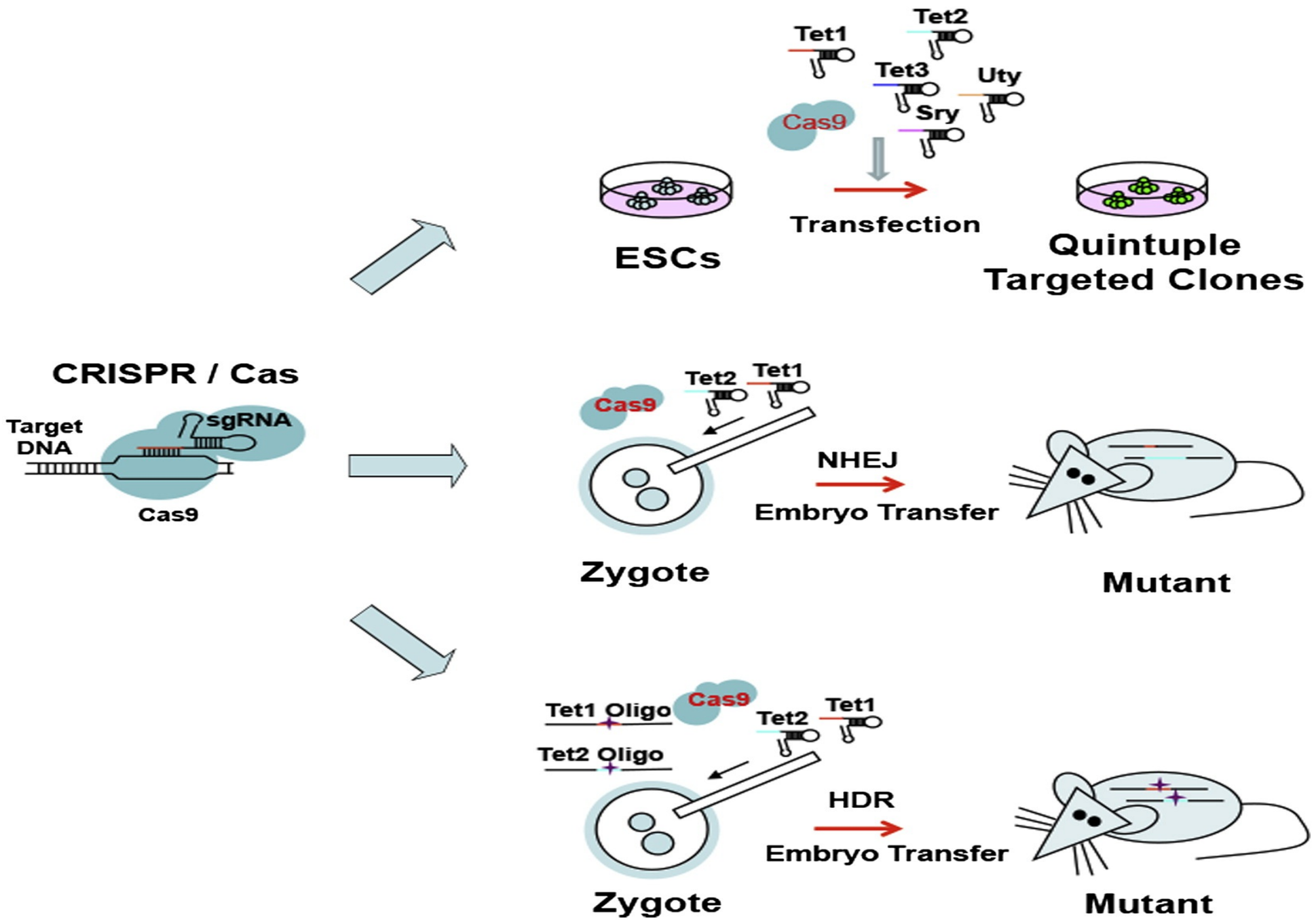
*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: gchurch@genetics.med.harvard.edu

Genome Editing



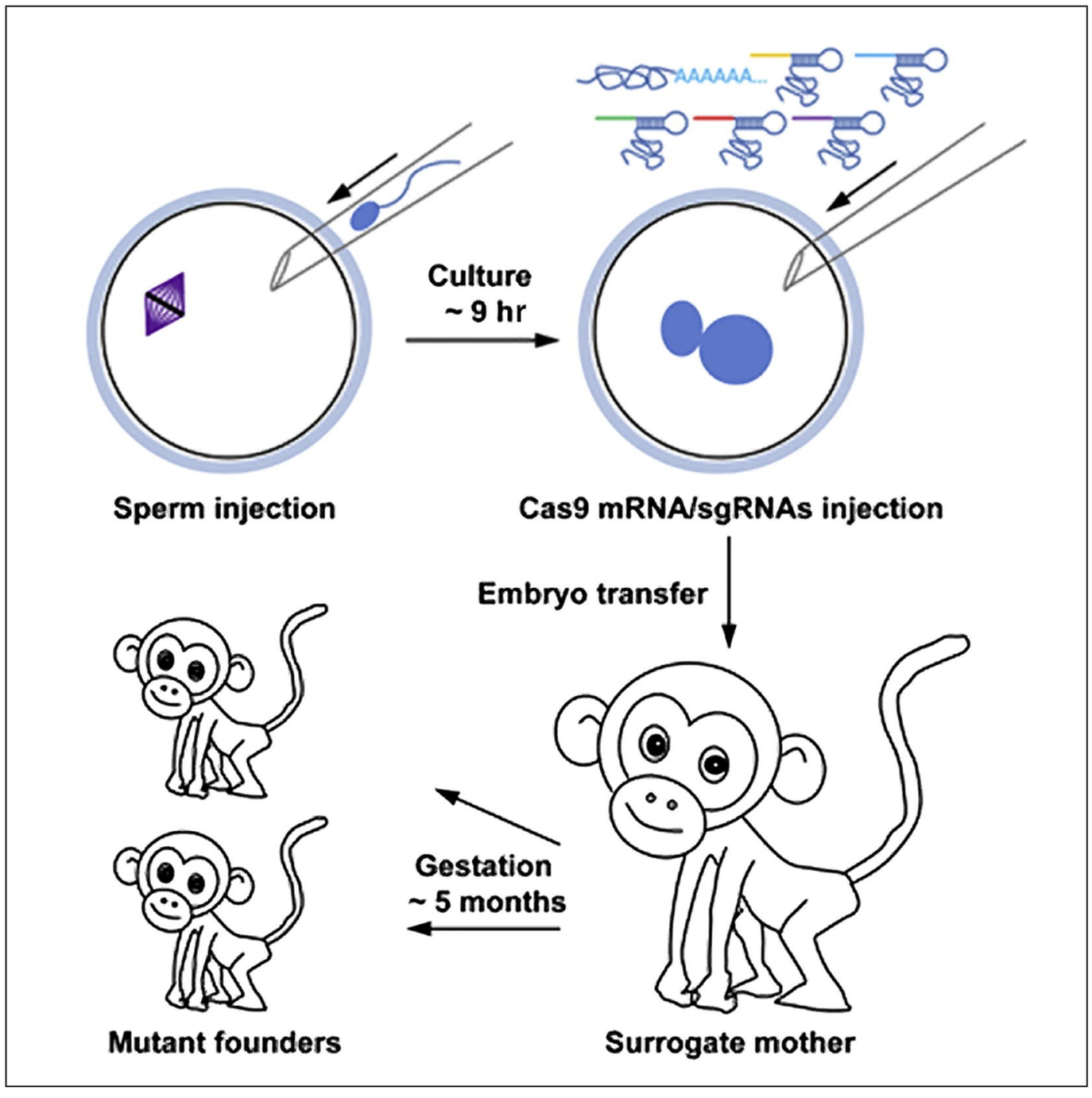
http://www.genome-engineering.org/crispr/

Genome Editing



Wang et al. *Cell* 2013

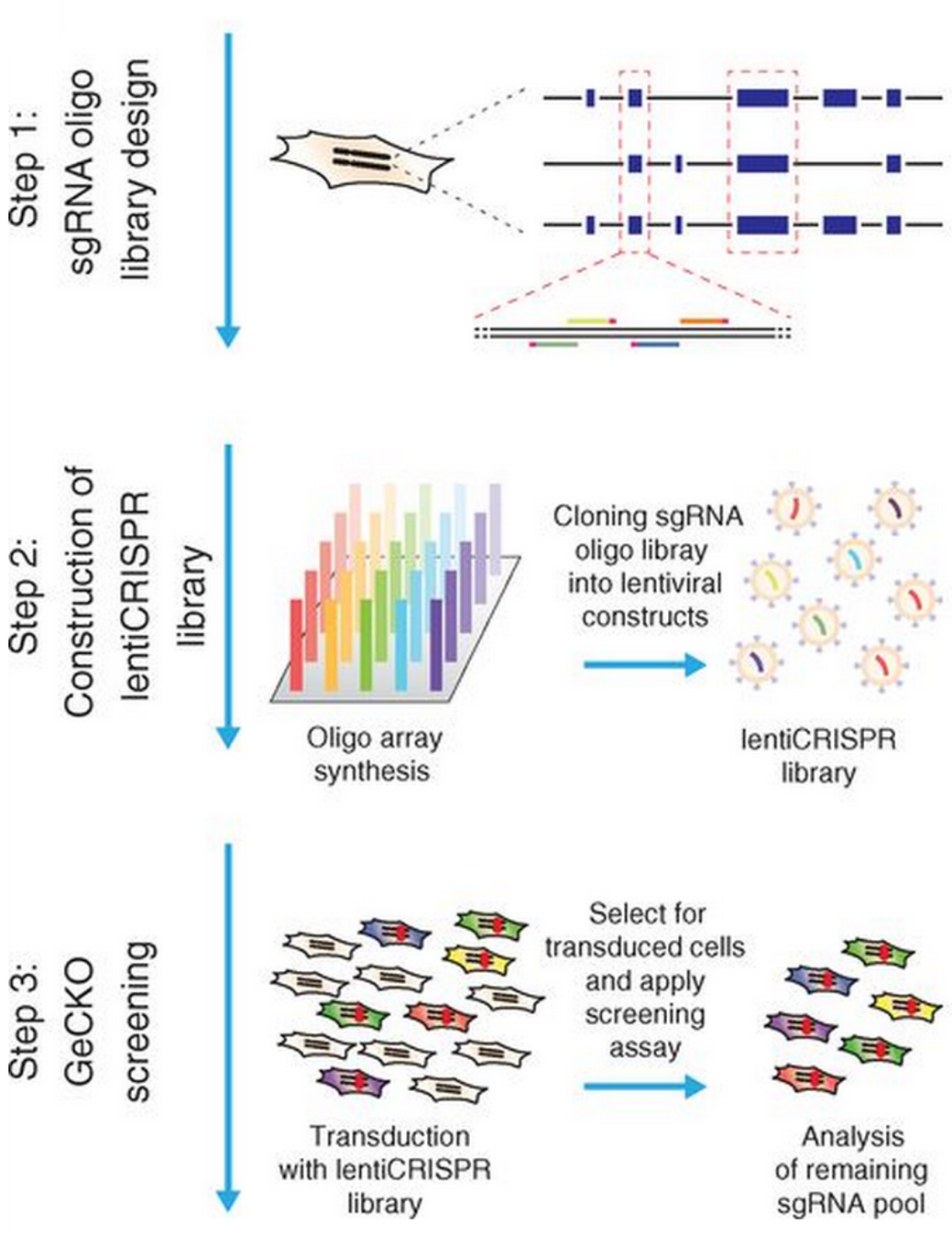
Primate Engineering

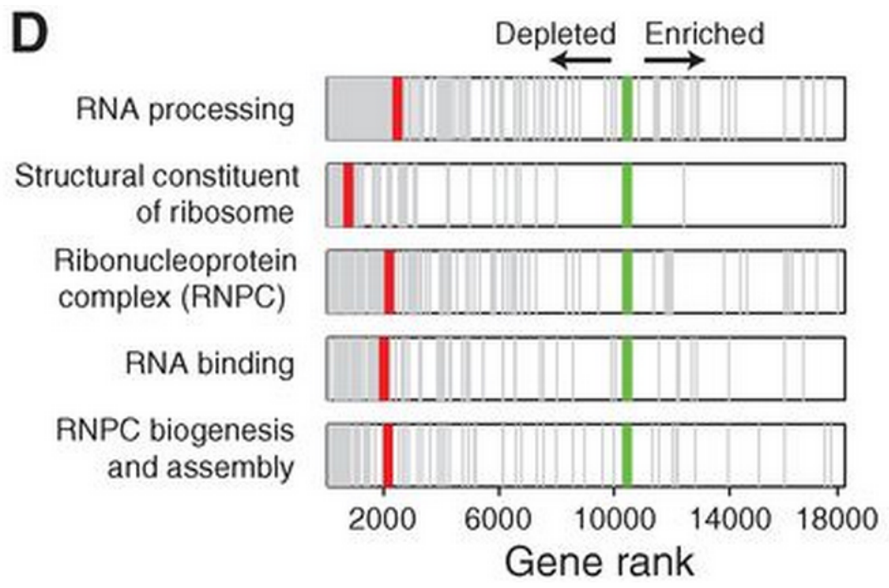


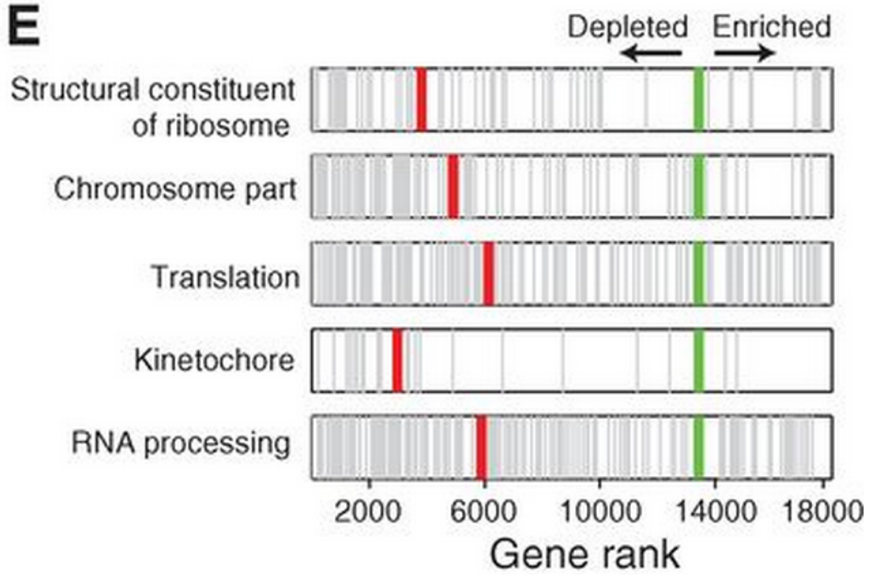
Niu et al. *Cell* 2014

Genome Screening

GeCKO
Genome-scale
CRISPR Knock-Out







Shalem et al. *Science* 2014

CRISPR/Cas Flexibility

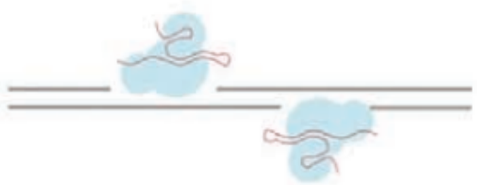
Cas9 nuclease or nickase cleavage

Cuts

Deletions

Nicks

Offset nicks

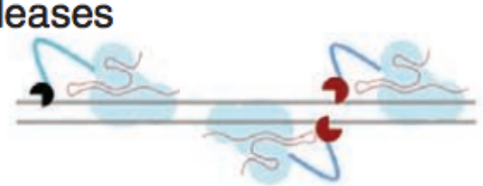


$Cas9_{nuclease\text{-}null}$ protein fusions

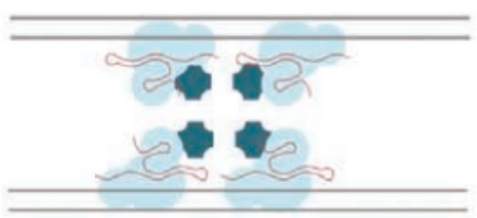
Regulation

Labeling

Nucleases



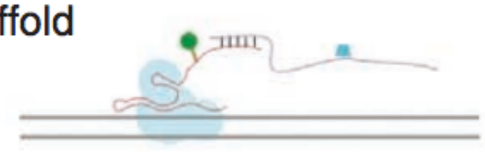
Recombinases



$Cas9_{nuclease\text{-}null}$ nucleic acid tethers

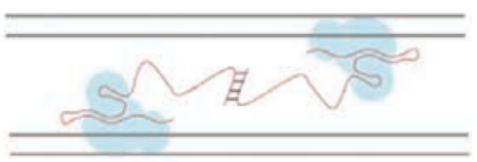
Regulation

Scaffold



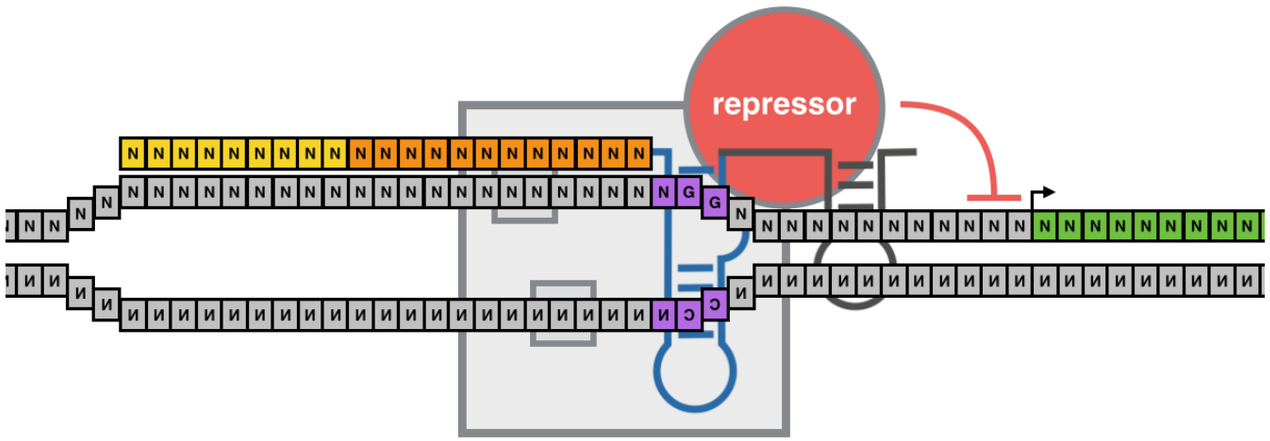
Competition

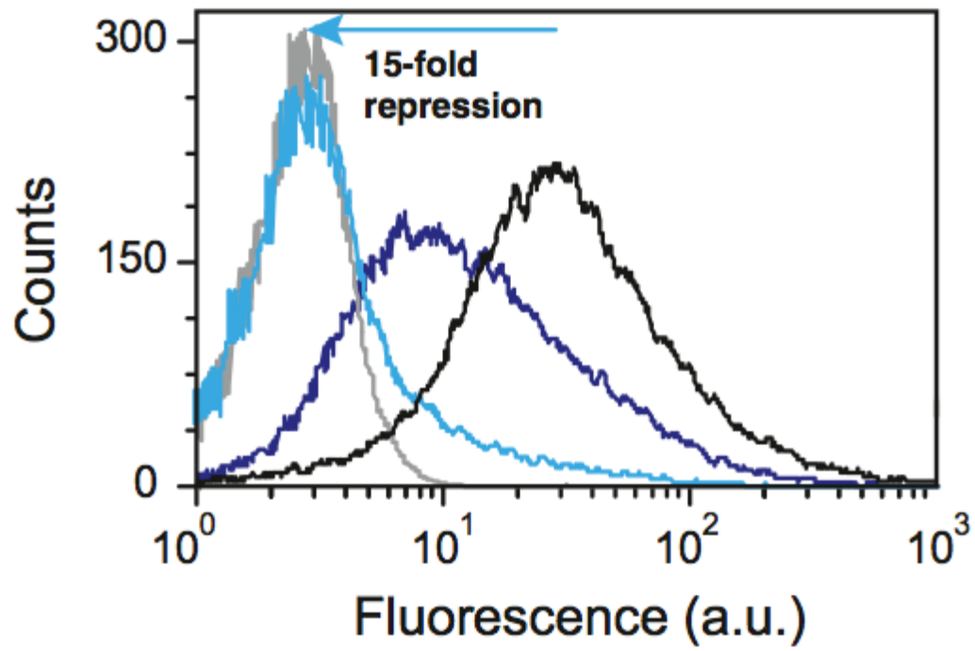
Aggregation

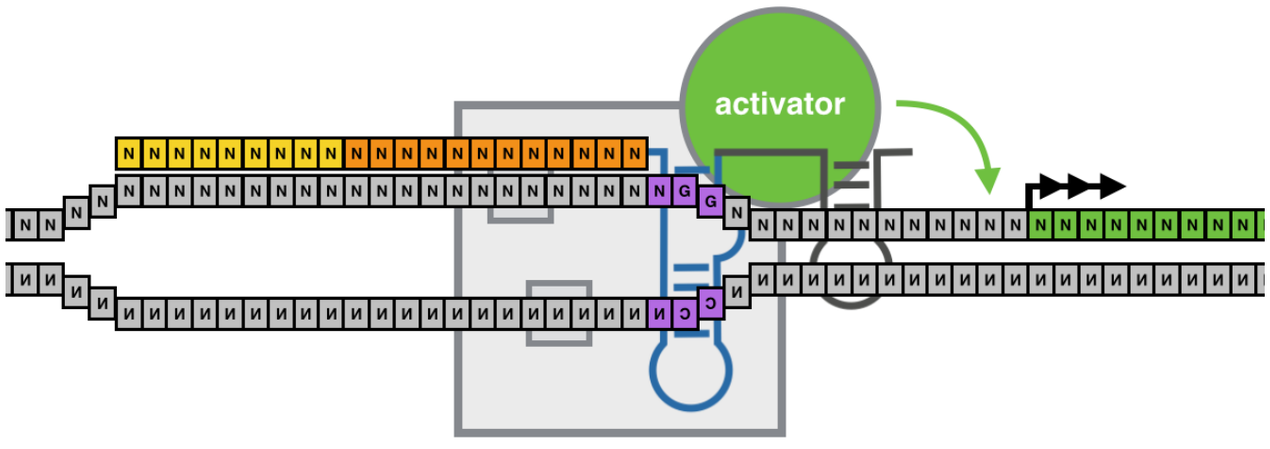


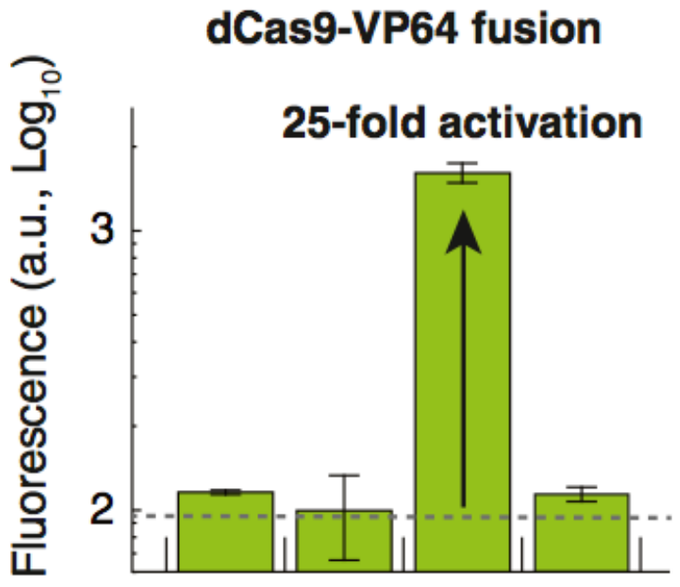
Mali et al. *Nature Methods* 2013

Genome Regulation



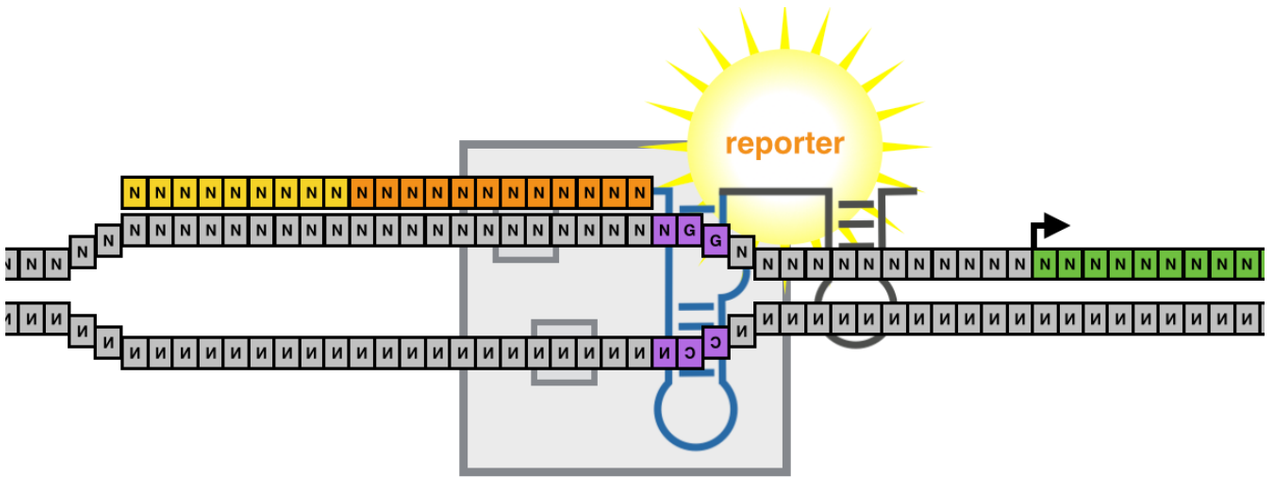


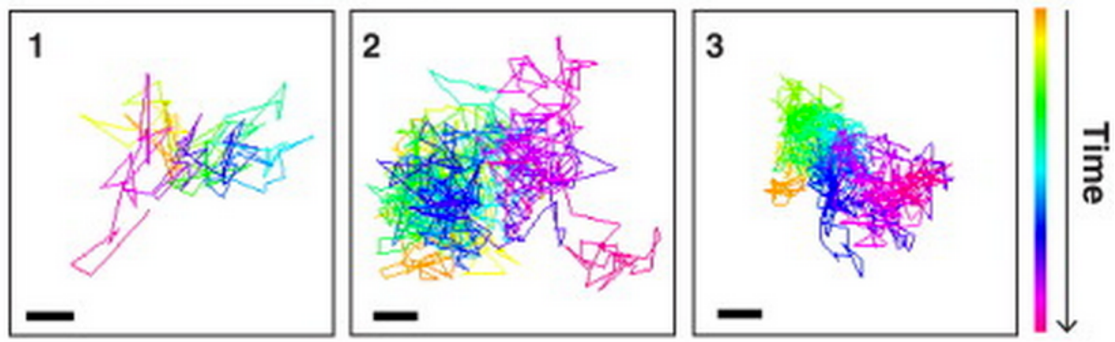


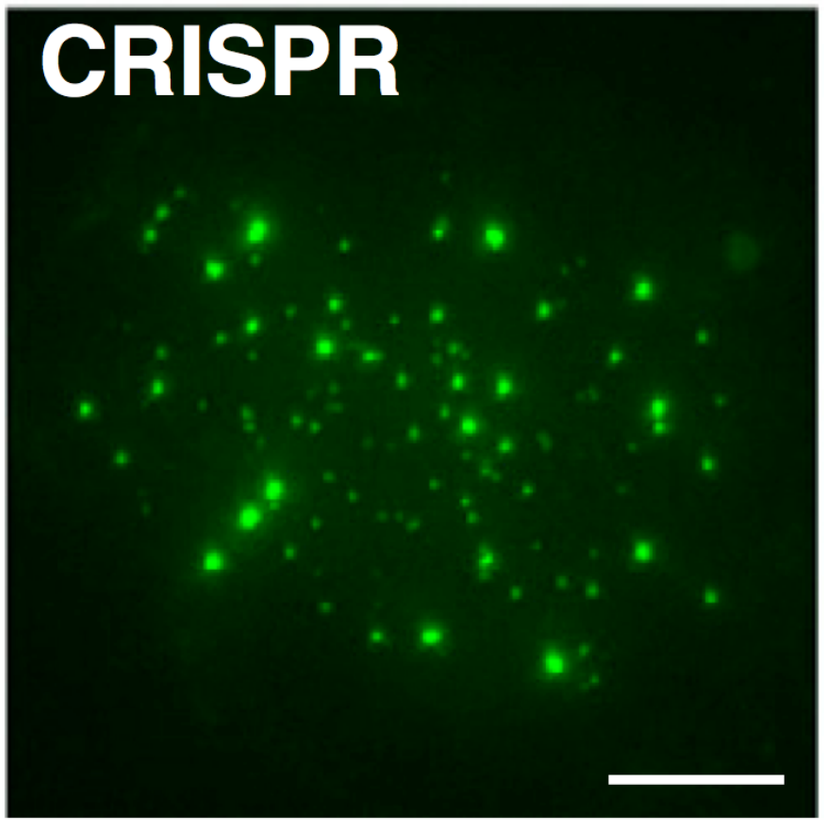


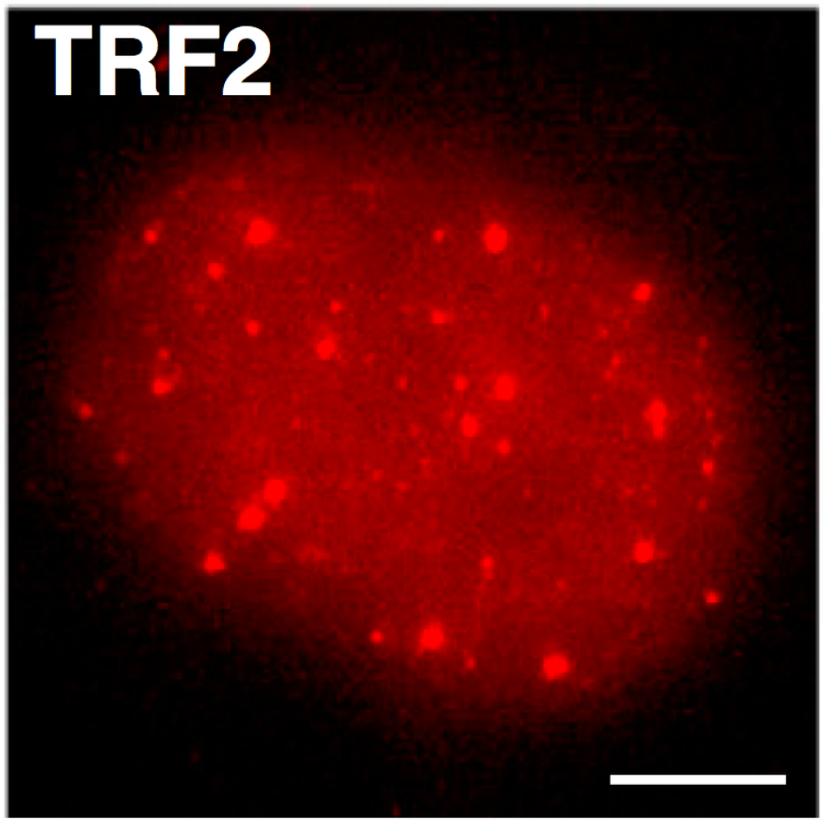
Genome Illumination

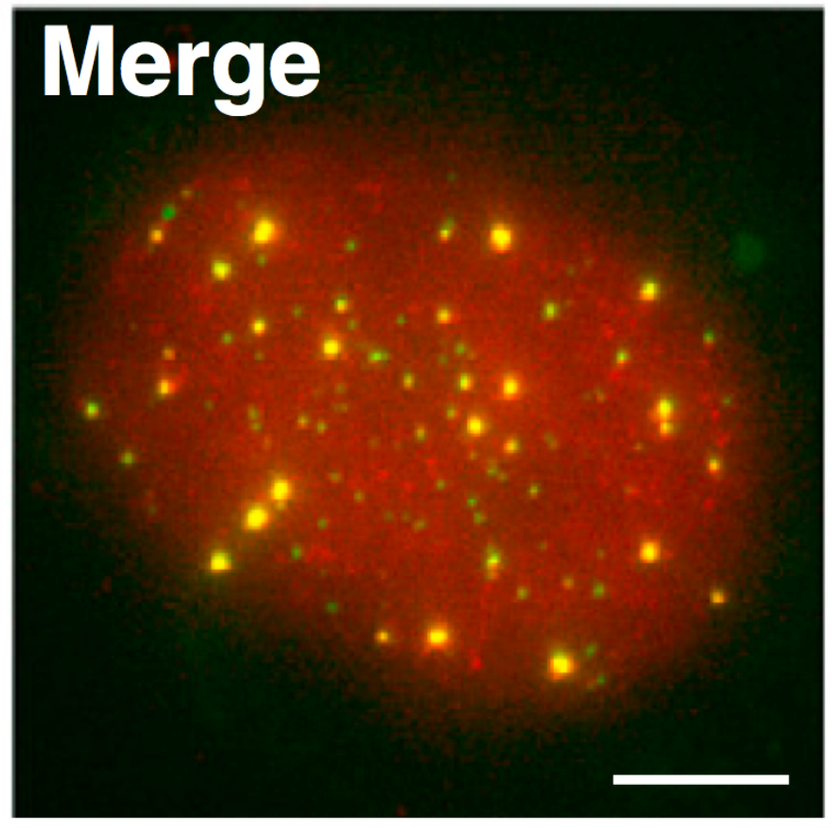
Cas9-Reporters allow for observation of genome dynamics.





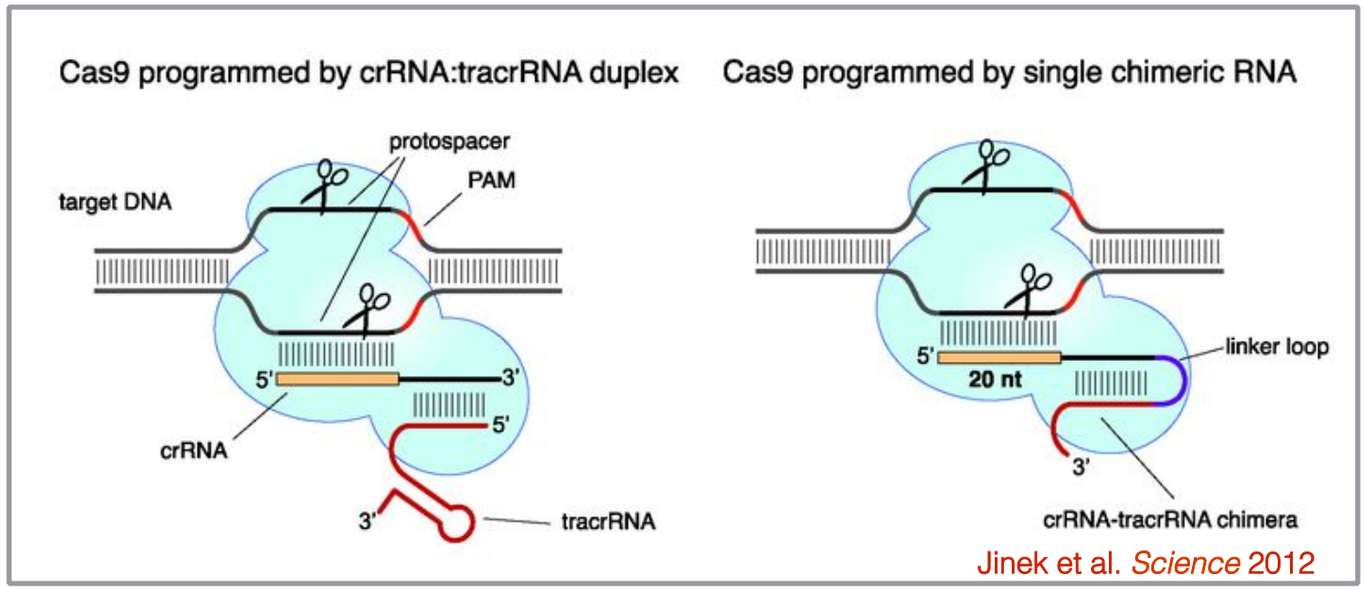




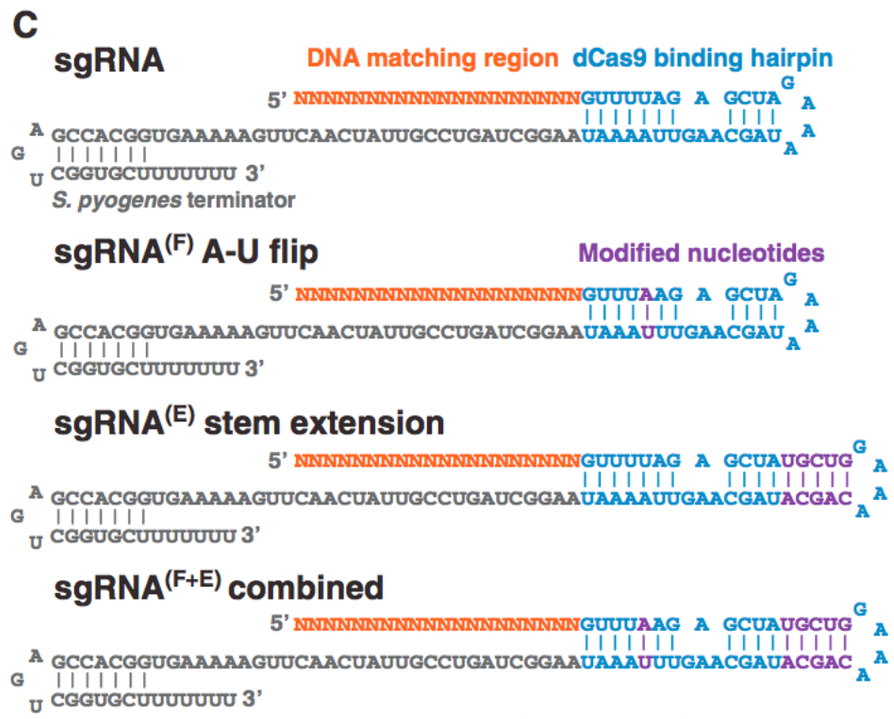


Chen et al. *Cell* 2014

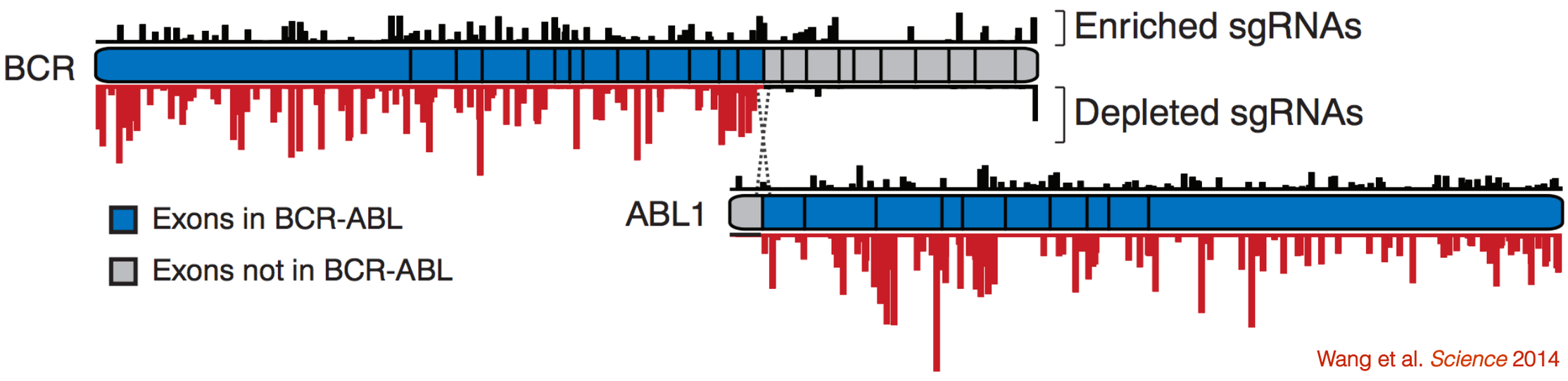
CRISPR Optimization



Jinek et al. *Science* 2012

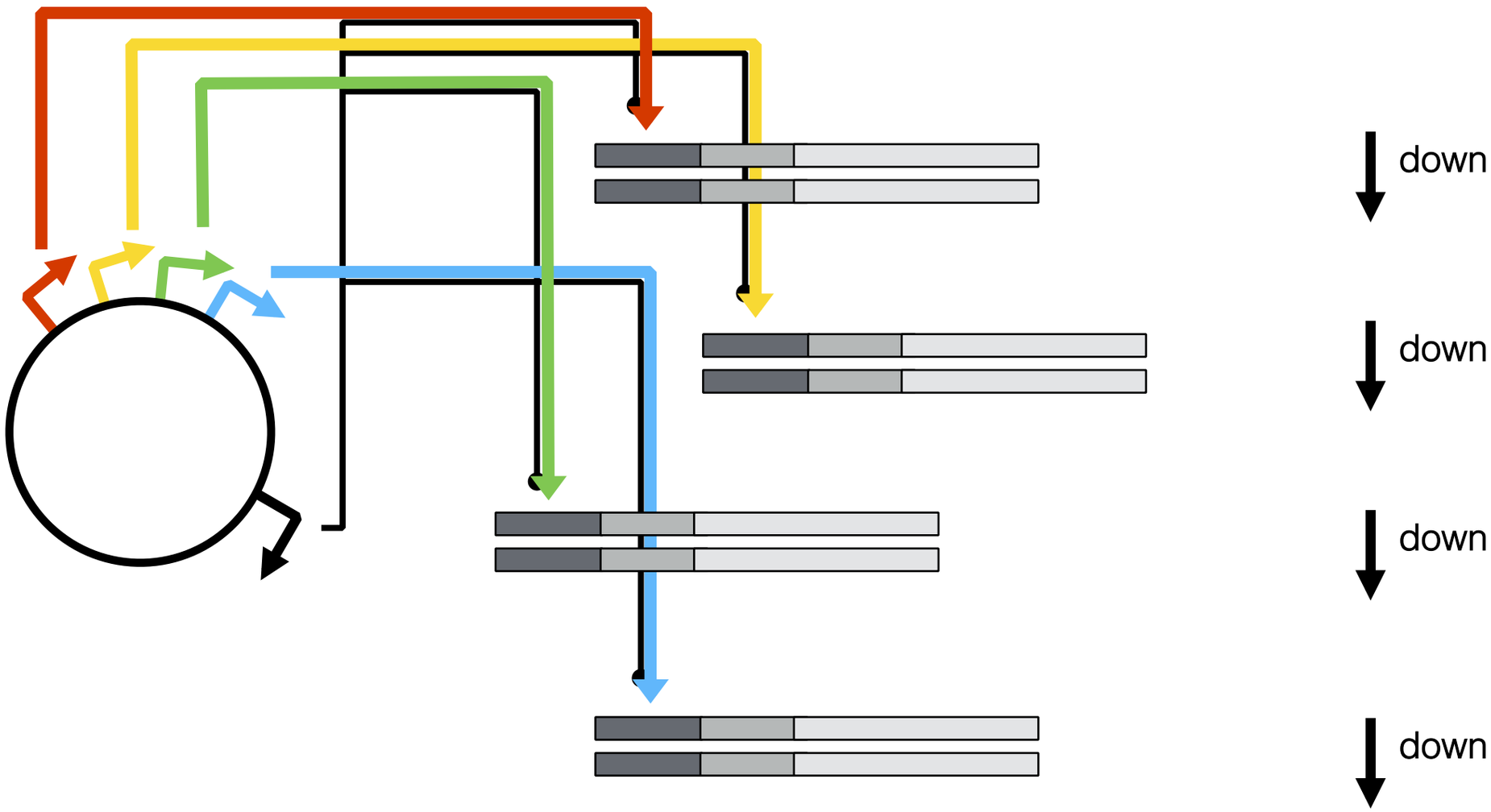


Chen et al. *Cell* 2014



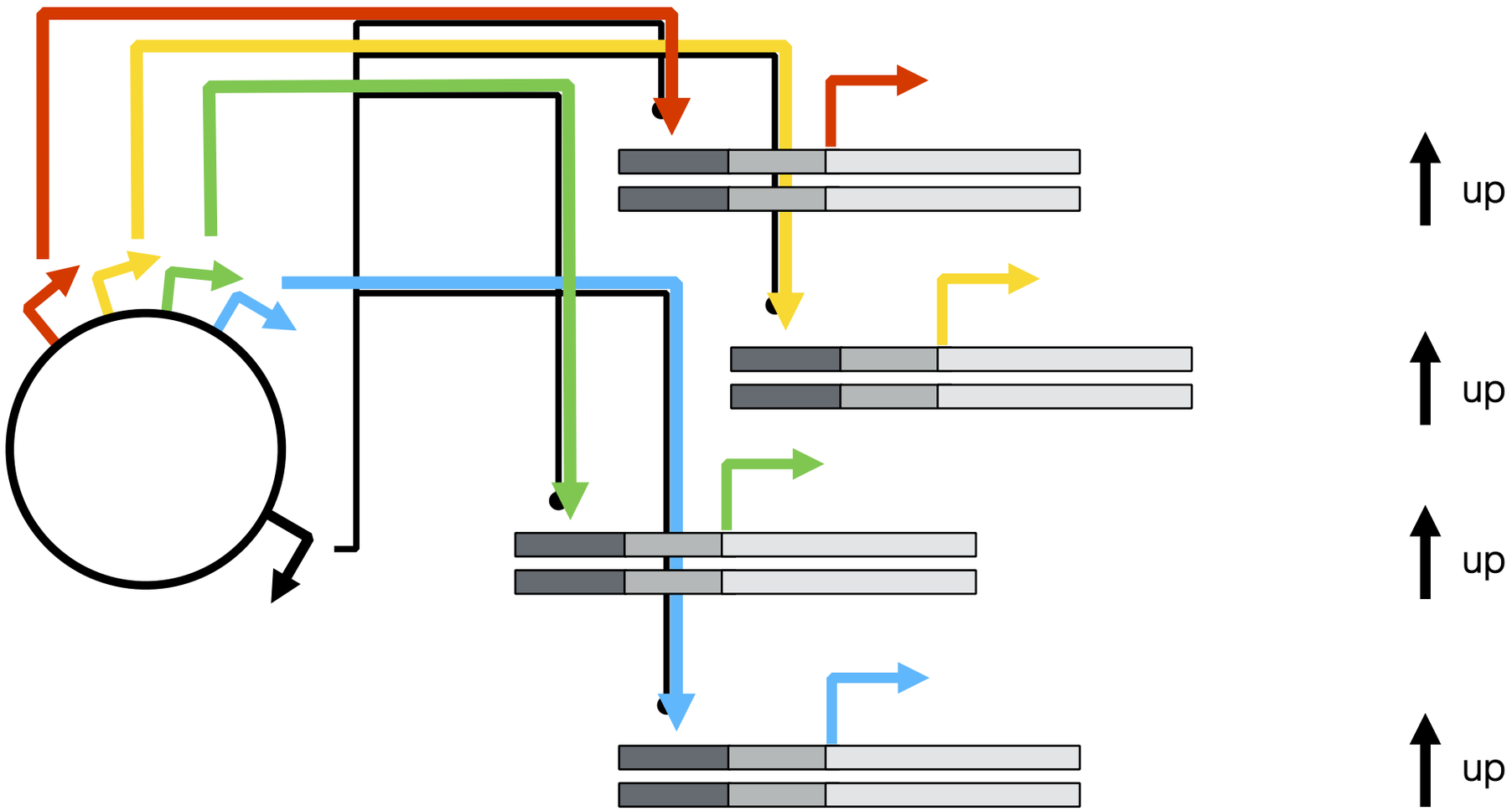
Wang et al. *Science* 2014

Multiplex Genome Regulation



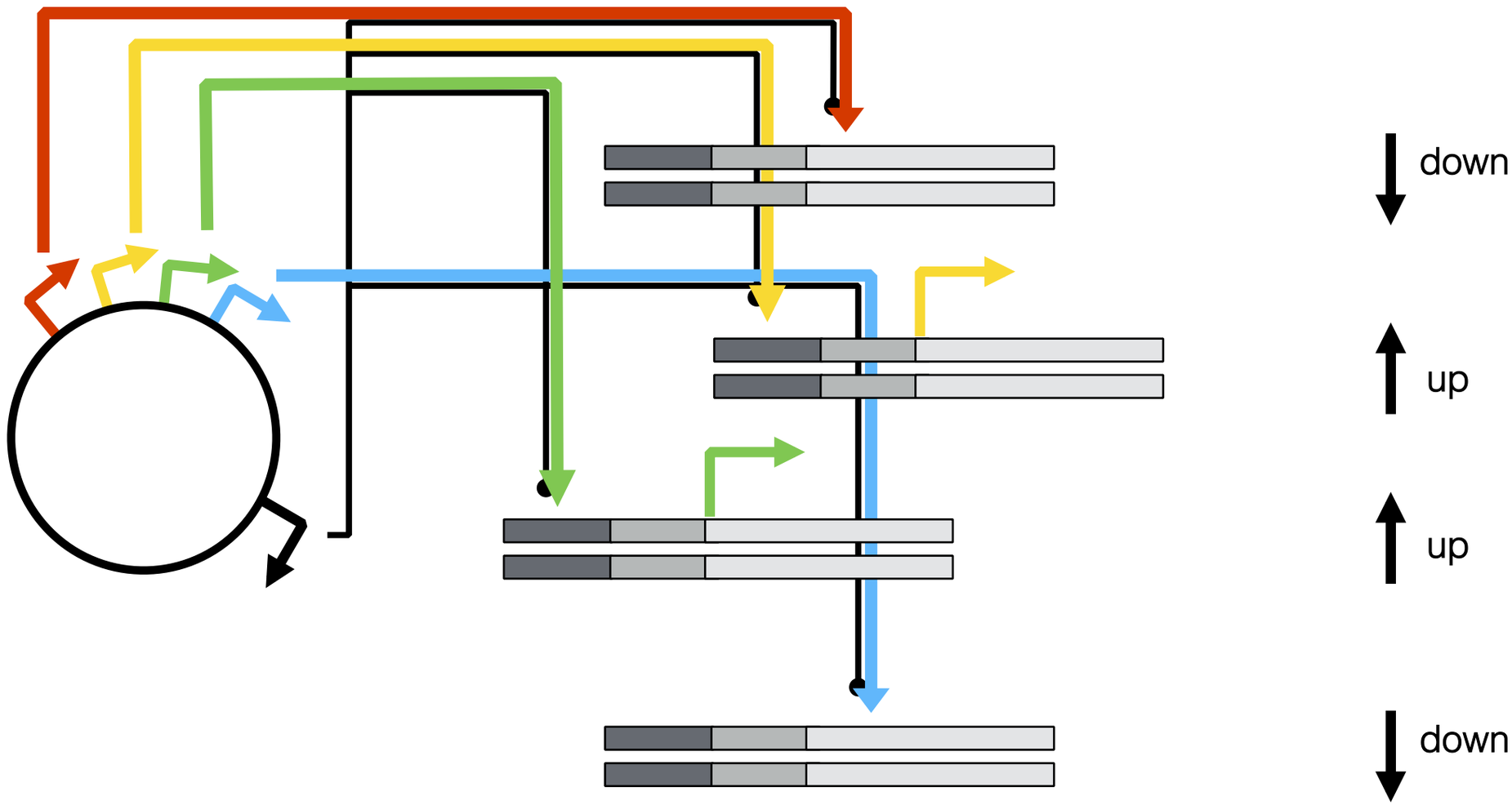
Multiple gRNAs with repressive Cas9

Multiplex Genome Engineering



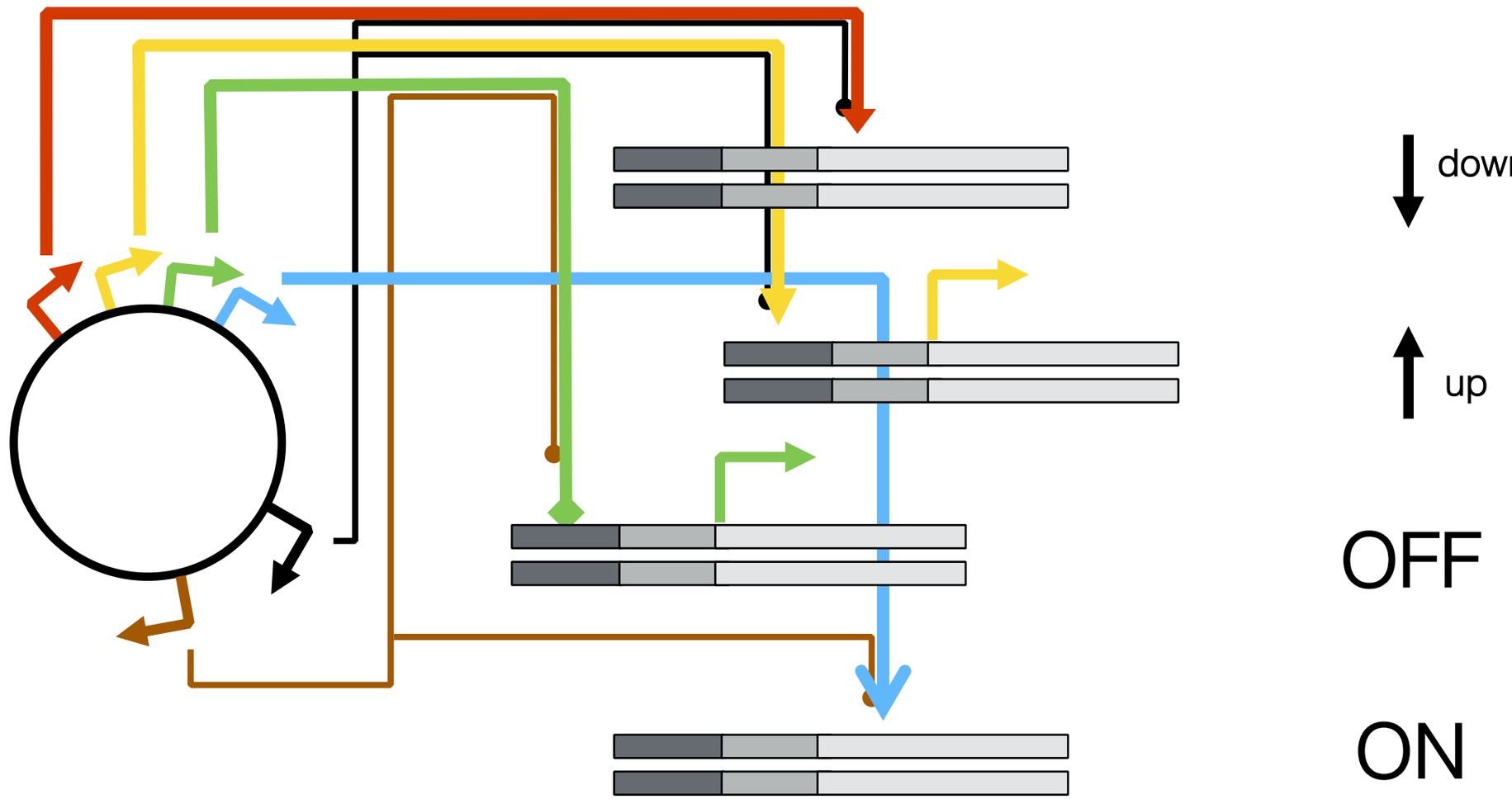
Multiple gRNAs with activator Cas9 targeting enhancer loci

Multiplex Genome Engineering



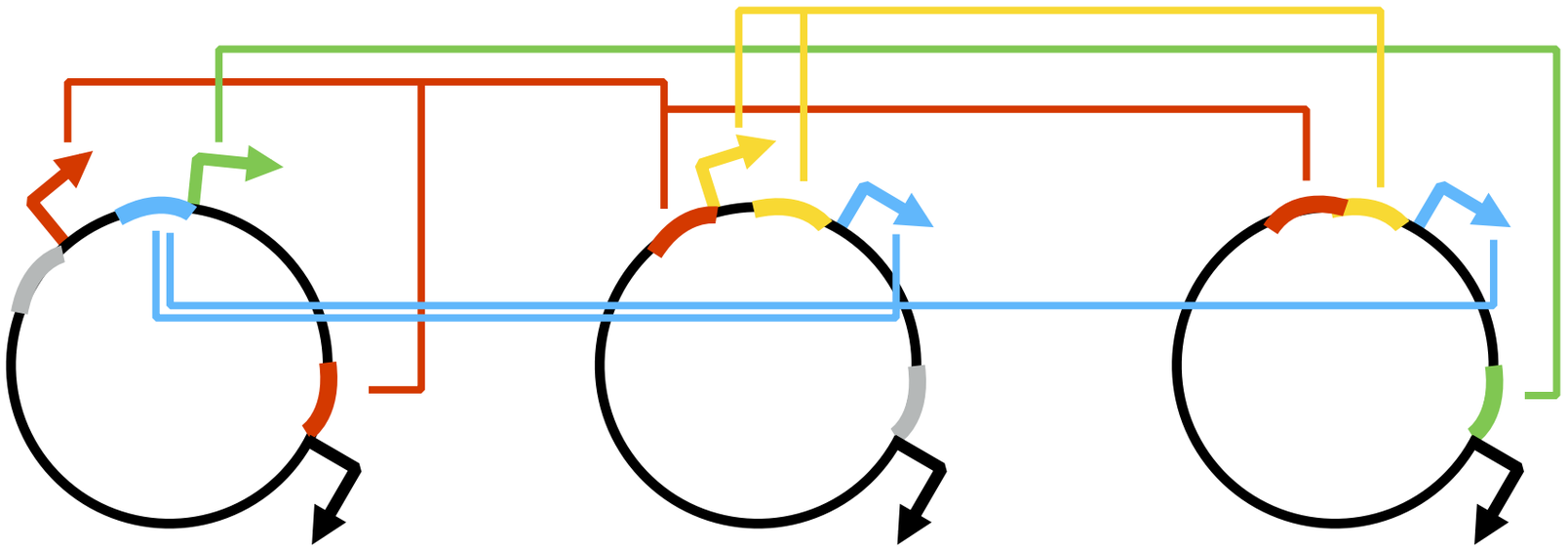
Multiple gRNAs with activator Cas9 targeting enhancer and suppressor loci

Multiplex Genome Engineering



Multiple gRNAs with orthogonal Cas9s and cognate gRNAs can modify and regulate genomes

CRISPR-based Genetic Circuits



gRNAs can regulate other gRNAs through conventional and synthetic promoters