Characterization of Cas9–Guide RNA Orthologs

Jonathan L. Braff,1,4 Stephanie J. Yaung,1,2,3,4 Kevin M. Esvelt,1 and George M. Church1,2,5

1Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, Massachusetts 02115; 2Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115; 3Program in Medical Engineering and Medical Physics, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

INTRODUCTION

As opposed to large multisubunit complexes of Type I and III CRISPR systems, a single Cas9 protein in the Type II system can carry out the prokaryotic adaptive immune response to foreign DNA. In native prokaryotic Type II CRISPR systems, transcribed arrays are processed into CRISPR RNAs (crRNAs) that form a complex with Cas9 and a trans-activating RNA (tracrRNA) (Deltcheva et al. 2011). The crRNA guides Cas9 to double-stranded DNA sequences called protospacers that match the sequence of the spacer and are flanked by a protospacer-adjacent motif (PAM) unique to the CRISPR system (Mojica et al. 2009). If spacer–protospacer base-pairing is a close match, Cas9 cuts both strands of DNA.

In June of 2012, Jinek et al. first described a functional chimera of crRNA and tracrRNA called a guide RNA or gRNA (Jinek et al. 2012). The introduction of the gRNA format simplified the Type II CRISPR system from four components (target, tracrRNA, crRNA, and Cas9) to three and sparked rapid technology development in genome engineering. In December of that year, simultaneous publications from the Church and Zhang laboratories showed the chimera as a tool for multiplexable genome engineering in human cells (Cong et al. 2013; Mali et al. 2013c).

To date, numerous Cas9-mediated functions and configurations for implementation have been presented. Besides the initial applications in targeted genome editing in bacteria (Jiang et al. 2013) and eukaryotes (Cong et al. 2013; Mali et al. 2013c), Cas9 has been used for programmable activation of
endogenous genes in bacteria (Bikard et al. 2012; Qi et al. 2012) and eukaryotes (Gilbert et al. 2013; Maeder et al. 2013; Mali et al. 2013a; Perez-Pinera et al. 2013). Cas9 can be used for epigenome regulation and visualization (Chen et al. 2013), as well as for targeted RNA degradation and localization with PAM-presenting DNA oligonucleotides (O’Connell et al. 2014). Although Cas9 naturally induces double-stranded DNA breaks, Cas9 variants can be engineered to cleave only a single strand (as a nickase) or lack endonucleolytic activity (as nuclease-null); these variants can be functionalized with other domains to recruit a variety of molecular components in a defined manner. In general, gRNAs can serve as a scaffold for assembling complexes of nucleic acids and proteins. One example is transcriptional initiation, for which there are various architectures for enhanced Cas9 activators (Chakraborty et al. 2014; Konermann et al. 2014; Chavez et al. 2015).

ORTHOGONALITY VALUE

Although powerful, a single Cas9 protein is able to mediate only one activity targeted at multiple different sites; it cannot carry out a different activity at other sites. Achieving simultaneous functionalities in a single cell requires multiple Cas9–gRNA orthologs, each engineered with a custom effector domain or modality to independently target its activity to its own array of target sites.

Given the vast potential of different Cas9-mediated functions (Mali et al. 2013b), orthogonal Cas9–gRNA pairs will be critical for independent targeting such that each function exclusively responds to its own set of gRNAs. This exclusivity insulates signal transduction pathways (Podgornaia and Laub 2013); there should be minimal, ideally zero, cross talk between Cas9–gRNA orthologs used in the same cell. Well-insulated, orthogonal biological control elements should vary independently without interfering with other components (Purnick and Weiss 2009). Orthogonal systems are further insulated by PAM specificity resulting in distinct sets of targetable protospacers. The properties of orthogonal Cas9 systems include the Cas9–gRNA interaction specificity as well as target recognition using distinct PAM sequences.

Identifying Putatively Orthogonal Cas9 Proteins

Putatively orthogonal Cas9 proteins can be selected by examining and identifying divergent repeat sequences. Tools like CRISPRFinder (Grissa et al. 2007a) and CRISPRdb (Grissa et al. 2007b) enable identification of CRISPR arrays with their constituent spacer and repeat sequences. There are also methods to experimentally validate expression and coprocessing of tracrRNAs and pre-crRNAs (Chylinski et al. 2013), and the dual RNA format can be engineered into the single gRNA format (Deltcheva et al. 2011). Candidate gRNA sequences can then be assessed to determine whether their sequences are sufficiently divergent to impart specificity to their cognate Cas9 protein. This is especially important because there is evidence for gRNA exchangeability among different Cas9s (Fonfara et al. 2013).

Characterizing Orthogonal Cas9 Proteins

PAMs are required for initial target binding, unwinding for interrogation, and subsequent cleavage of target sequences (Anders et al. 2014; Sternberg et al. 2014). PAMs are generally unique to their cognate Cas9–gRNA pair and can be identified in three ways. First, using bioinformatics tools, a multiple sequence alignment of targeted bacteriophages or plasmids can yield a consensus PAM. However, this method requires searching available sequences of phages and plasmids for matches to CRISPR spacers. It is thereby unable to predict PAMs with confidence if the availability of relevant sequences is insufficient. In addition, the bioinformatics approach may be affected by biases imparted by endogenous spacer acquisition machinery that is not generally used in engineering applications. Experimentally, in vitro cleavage assays with different plasmid substrates can produce a position weight matrix of potential PAMs, but this is relatively low-throughput compared to a library approach (Esvelt et al. 2013), which can interrogate a PAM sequence space of NNNNNNNNN (corresponding to \(4^6 = 65,536\) sequences). In theory, one could perform a library selection on \(4^{10} = 1,048,576\) sequences,
given enough sequencing depth, if the PAM is suspected to extend out to 10 bases. For more details on this approach, see Protocol: Characterizing Cas9 Protospacer-Adjacent Motifs with High-Throughput Sequencing of Library Depletion Experiments (Braff et al. 2016).

Because the PAM sequence (Table 1) is key to distinguishing “self” versus “nonself” DNA, characterizing the PAM is important not only for Cas9-mediated activity, but also for understanding potential off-targets (Aach et al. 2014; Zhang et al. 2014). Thus, it is an essential first step to determining the orthogonality of CRISPR systems. Many groups have studied ways to improve Cas9 specificity, as Cas9–gRNA complexes are known to tolerate up to three mismatches in their targets. Notably, specificity should be evaluated with the context of the assay in mind; there are observed differences between cleavage assay results and in vivo performance in terms of PAM preference (Fonfara et al. 2013), and off-target binding does not necessarily result in off-target cleavage (Cencic et al. 2014; Wu et al. 2014). Specificity also depends on the concentration and bioavailability of the components. Factors affecting specificity include the amount of Cas9 protein, gRNA expression, and half-life of components, as well as the target and PAM sequence accessibility (Fu et al. 2013; Hsu et al. 2013; Mali et al. 2013a; Pattanayak et al. 2013). Some approaches to increase specificity include using multiple homologs for activity, finding or engineering improved variants, harnessing the dimerization requirement of FokI (Guilinger et al. 2014; Tsai et al. 2014), and using slightly truncated gRNAs (Fu et al. 2014).

**CONCLUSION**

Expanding the set of characterized Cas9–gRNA orthologs increases the number of engineered functions that can be simultaneously deployed in a single cell. Investigating additional Cas9 proteins can also provide insights into the nuances of varying sensitivity, specificity, and kinetic requirements in the context of the engineering application. A suite of Cas9 homologs from which to choose would allow careful matching and optimization of these parameters for a desired application. Further, the differences between PAMs found in Cas9 homologs increase the range of targetable sequences.

Undoubtedly, Type II CRISPR systems have been widely adopted for their multiplexability and programmability. The focus has been on the performance of a single function at a large number of loci and the ability to rapidly retarget this function. The increasing number of engineered variants capable of performing distinct functions has made the ability to program multiple target action pairs a very promising avenue for technology advancement.

**ACKNOWLEDGMENTS**

This work was supported by U.S. Department of Energy grant DE-FG02-02ER63445 (to G.M.C.) and the Wyss Institute for Biologically Inspired Engineering.
Characterization of Cas9–gRNA Orthologs

REFERENCES


Cite this introduction as *Cold Spring Harb Protoc*; doi:10.1101/pdb.top086793

425
Characterization of Cas9–Guide RNA Orthologs

Jonathan L. Braff, Stephanie J. Yaung, Kevin M. Esvelt and George M. Church

Cold Spring Harb Protoc; doi: 10.1101/pdb.top086793

Email Alerting Service
Receive free email alerts when new articles cite this article - click here.

Subject Categories
Browse articles on similar topics from Cold Spring Harbor Protocols.

- Bioinformatics/Genomics, general (180 articles)
- Characterization of Protein Complexes (80 articles)
- Computational Biology (100 articles)
- Molecular Biology, general (1104 articles)
- Proteins and Proteomics, general (533 articles)
- Proteomics (64 articles)
- RNA (254 articles)
- RNA, general (220 articles)