Applications of Cas9 as an RNA-programmed RNA-binding protein

*David A. Nelles[1)], Mark Y. Fang[1)], Stefan Aigner[1)] and Gene W. Yeo[1)2)]**

The *Streptococcus pyogenes* CRISPR-Cas system has gained widespread application as a genome editing and gene regulation tool as simultaneous cellular delivery of the Cas9 protein and guide RNAs enables recognition of specific DNA sequences. The recent discovery that Cas9 can also bind and cleave RNA in an RNA-programmable manner indicates the potential utility of this system as a universal nucleic acid-recognition technology. RNA-targeted Cas9 (RCas9) could allow identification and manipulation of RNA substrates in live cells, empowering the study of cellular gene expression, and could ultimately spawn patient- and disease-specific diagnostic and therapeutic tools. Here we describe the development of RCas9 and compare it to previous methods for RNA targeting, including engineered RNA-binding proteins and other types of CRISPR-Cas systems. We discuss potential uses ranging from live imaging of transcriptional dynamics to patient-specific therapies and applications in synthetic biology.

Keywords:
Cas9; CRISPR-Cas; RCas9; RNA-binding proteins; RNA biology; RNA targeting

DOI 10.1002/bies.201500001

[1)] Department of Cellular and Molecular Medicine, Stem Cell Program, and Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA
[2)] Molecular Engineering Laboratory, Biomedical Sciences Institutes, Agency for Science, Technology & Research and Yong Loo Lin School of Medicine, National University of Singapore, Singapore

***Corresponding author:**
Gene W. Yeo
E-mail: geneyeo@ucsd.edu

Abbreviations:
CRISPR, clustered regularly interspaced short palindromic repeats; **GFP,** green fluorescent protein; **GOI,** gene of interest; **KH,** K homology; **lncRNA,** long non-coding RNA; **MCP,** MS2 coat protein; **ORF,** open reading frame; **PAM,** protospacer adjacent motif; **PUF,** pumilio and FBF (Fas binding factor) homology protein; **RBP,** RNA-binding protein; **RRM,** RNA recognition motif; **sgRNA,** single-guide RNA; **siRNA,** short interfering RNA; **UTR,** untranslated region.

Introduction

The human genome project was completed more than a decade ago and sets the foundation for understanding the genetic basis of cell behavior in health and disease. Since then, efforts have shifted towards understanding the importance of functional genetic elements and how they affect gene expression [1]. Since all cells of an individual contain largely the same DNA, the functional distinctions between cell types (a cardiomyocyte and a neuron, for instance) are closely linked to the portions of the genome that are transcriptionally active. As a result, measurement of transcribed RNA within individual cells reveals cellular identity and distinguishes healthy and disease states. For example, expression levels of a focused panel of RNA transcripts identified disease-associated aberrations in neuronal development in models of autism spectrum disorder [2]. As another example, the expression of certain small non-coding RNAs known as microRNAs (miRs) is increasingly recognized as a characteristic signature of oncogenic transformation. Tumor microRNA signatures can serve as biomarkers informing the type of malignancy and associated clinical outcomes [3, 4]. These studies and others make clear that tracking informative RNAs in vivo will be key to disease modeling, diagnostics and potentially therapeutics.

Due to the obvious impact of expressing specific RNAs on cell state and behavior, unraveling the mechanisms that affect the processing of these RNA has become very important. Following transcription, protein-encoding RNAs undergo a series of maturation steps that include alternative splicing, nuclear export, and subcellular targeting, turnover, and spatiotemporally restricted translation. These steps are mediated by RNA-binding proteins (RBPs) and dysfunction of these factors and their RNA targets causes disease in humans [5]. Altered subcellular distribution of RBPs caused by gain-of-function expanded RNA elements is also becoming a common theme in human disease. For example, expansion of an intronic hexanucleotide repeat within the *C9ORF72* gene was recently recognized as the most frequently mutated genetic locus among two common neurodegenerative disorders, frontotemporal lobar degeneration and amyotrophic lateral sclerosis [6, 7]. In vivo approaches to targeting the processing of endogenous RNA would open up basic

biological understanding of development and disease as well as new avenues for therapies.

A recent publication has raised awareness of the potential of RNA-guided RNA recognition [8]. Here, we focus on the potential of repurposing Cas9, the effector nuclease of the *Streptcococcus pyogenes* CRISPR-Cas system that has been used to recognize DNA in mammalian cells, as an RNA-programmed RNA recognition technology.

Current RNA recognition modalities and their limitations

The development of designer RNA recognition factors will support a variety of advances in biology and medicine. Aside from targeted modulation of RNA processing and abundance, a designer RBP could generate completely novel activities in response to RNA recognition, such as generating a signal for noninvasive detection of cell state, promoting association of signaling proteins and their substrates only in particular cell types, or even ablating cells that display particular expression profiles. This broad potential has motivated the development of designer RNA recognition factors to varying degrees of success.

An ideal RNA recognition system would be capable of strong and specific binding to endogenous RNAs and display sufficient modularity for simple and predictable targeting. Inroads towards programmable RNA recognition have emerged based upon engineered natural nucleic acid binding proteins that are powerful for some applications but suffer from limited programmability, recognize too short a recognition sequence to be specific, and/or require large libraries of protein repeat sequences to target all possible RNA sequences. In contrast to direct recognition of nucleic acids by proteins, CRISPR-Cas (clustered regularly-interspaced short palindromic repeats) systems form bacterial adaptive immune systems and recognize invading nucleic acids with RNA-guided proteins.

An obvious strategy is the alteration or concatenation of natural RNA-binding protein domains. The identification of canonical RNA recognition protein domains such as KH and RRM led to attempts at identifying and modulating their natural RNA targets [9–11]. These domains bind RNA in groups of 4–5 contiguous nucleotides. As a result, libraries of more than 1,000 protein domains are required to recognize all 5-base RNA sequences. In contrast, PUF proteins contain repeat domains that recognize a single RNA nucleotide each so only four repeats are in principle required to recognize all possible RNA sequences. The crystal structures of natural PUF proteins were first described in 2001 [12] and revealed recognition of specific RNA bases that is largely determined by the amino acid side chains rather than the backbone. Since their initial discovery, the RNA specificity of PUF proteins has been decoded [13] and PUFs have been designed against a variety of RNA targets [14]. Furthermore, PUFs have been successfully fused to nucleolytic domains to target and destroy disease-associated RNA [15]. However, PUF proteins can only recognize eight contiguous bases and local secondary structures can have a strong influence on RNA affinity, thus limiting their utility [15].

Cas9 for RNA-guided nucleic acid recognition

While PUF, KH, and RRM proteins rely upon protein-RNA interactions to recognize RNA, nucleic acid base-pairing represents a simpler means of RNA recognition. The CRISPR-Cas bacterial immune system utilizes RNA-mediated base-pairing to recognize DNA, and has been successfully repurposed to target DNA in mammalian cells [16–19]. In bacteria and archaea, CRISPR-Cas forms the functional core of adaptive immune systems that are typically composed of a nuclease associated with a pair of RNAs called the trans-activating CRISPR RNA (tracrRNA) and CRIPSR RNA (crRNA). The tracrRNA and crRNA guide the CRISPR nuclease to invading plasmid or bacteriophage DNA by base-pairing for cleavage by the nuclease (Fig. 1A). Recently, a Type II CRISPR-Cas system from *S. pyogenes* was repurposed to target mammalian DNA by creation of an artificial combination of the tracrRNA and crRNA called the single guide RNA (sgRNA) [16, 17]. By allowing facile DNA targeting via the sgRNA sequence, RNA-programmed Cas9 is rapidly proving to be a popular means of genome editing and transcription modulation. The recent application of Cas9 to RNA targeting may support a similar shift in programmable RNA recognition based on RNA programming over engineered binding proteins.

RNA-targeted Cas9 (RCas9) is the subject of recent work from the Doudna lab that demonstrates strong and specific binding and subsequent cleavage of ssRNA by Cas9 in vitro [8]. In Figure 1, we compare this new approach to RNA recognition by Cas9 to DNA recognition. DNA targeting by Cas9 requires two features: an NGG sequence (where N = any nucleotide) referred to as the protospacer adjacent motif (PAM) and a sgRNA carrying an antisense sequence adjacent to the PAM (Fig. 1A). These two features are also required for RNA targeting by Cas9, although the PAM motif is provided by a hybridized antisense oligonucleotide (the PAMmer), which sits adjacent to the sgRNA antisense sequence after hybridization to the target RNA (Fig. 1B).

O'Connell and Oakes et al. [8] also demonstrated that a 5′ extension of the PAMmer beyond the PAM motif is required to generate specific RNA recognition programmed by the sgRNA. Shorter PAMmers lacking this extension promote Cas9:sgRNA binding that is independent of sgRNA sequence, but the sequence specificity of sgRNA-programmed RNA recognition is reconstituted by an extension of the PAMmer. This effect may be due to the energetic cost of Cas9-mediated unwinding of the PAMmer-RNA target duplex which is recovered only when the sgRNA hybridizes its target. Since the sgRNA is encodable and small (~100 bases), there is potential to generate large libraries of sgRNAs to target particular gene networks or screen the transcriptome. In contrast, the size of engineered RNA recognition proteins does not easily support large-scale screens. Although the cost associated with producing and distributing large libraries of modified oligonucleotide PAMmers will be an obstacle to work at this scale, future developments may allow the use of minimally modified oligonucleotides and leverage low-cost, high-throughput oligonucleotide synthesis technologies.

The aforementioned study was conducted exclusively in vitro and the strength and specificity of RNA-targeting Cas9

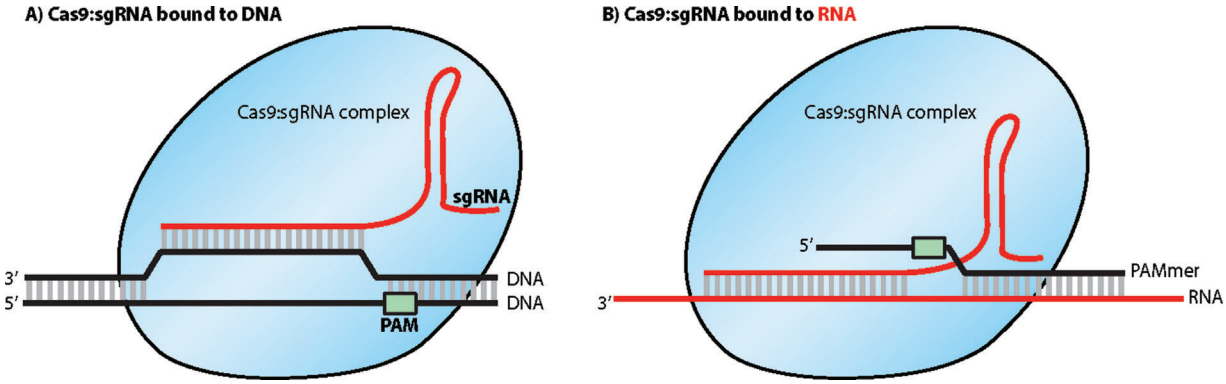


Figure 1. *S. pyogenes* Cas9 and sgRNA complexes bound to DNA or RNA. **A:** The Cas9:sgRNA complex requires a DNA NGG motif referred to as the protospacer adjacent motif (PAM). In the case of DNA binding, the PAM is supplied by the DNA target itself. The mechanism of DNA targeting by Cas9 is described extensively elsewhere. **B:** RNA-targeted Cas9 (RCas9) relies upon a short oligonucleotide called the PAMmer to supply the PAM motif. By utilizing a mismatched PAMmer, specificity of RCas9 for RNA while avoiding the encoding DNA is achieved. The PAMmer also carries a 5′ overhang which is required to maintain target specificity conferred by the sgRNA. As a result, it is hypothesized that the 5′ end of the PAMmer is at least partially dehybridized from the target RNA as Cas9-mediated unwinding of the PAMmer:target RNA duplex may confer an energetic cost that is recovered when the sgRNA hybridizes the target RNA.

(RCas9) inside living cells or organisms is not yet known. Analogous to recent measurement of CRISPR-Cas off-target activities on genomic DNA [18, 19], extensive validation of the RCas9 binding specificity will be required in order to evaluate its potential as a intracellular, RNA-programmable RNA-binding protein. Along with its well-known ability to target DNA, the comprehensive ability of the *S. pyogenes* CRISPR-Cas system to target nucleic acids is now being established. Box 1 highlights major challenges that must be overcome for RCas9 to be applicable in vivo.

Modulating post-transcriptional gene expression

A natural application of RCas9 utilizes the inherent endonucleolytic activity of Cas9 to attenuate gene expression via cleavage of particular transcripts (Fig. 2A). While RNA interference (RNAi) supports effective RNA recognition and cleavage, RCas9-based gene knockdown could be particularly

Box 1

In vivo applications of RCas9

An evaluation of the potential of RCas9 for RNA targeting in living organisms naturally begins by examining reported in vivo applications of Cas9 for genome editing. Delivery of Cas9 and the cognate sgRNA have been achieved by various means, including the use of viruses that encode Cas9 and the sgRNA [51, 52], transgenic animals that allow drug-inducible expression of Cas9 [53], and delivery of Cas9 protein and sgRNA via anionic fusion proteins and cationic lipids [54]. Modulation of RNA splicing by RCas9 via targeting of a splicing factor fused to Cas9 to a pre-mRNA of interest, for instance, could be conducted in the central nervous system with an appropriately serotyped adenovirus. Splicing modulation in other tissues could be achieved with drug-inducible and tissue-specific expression of Cas9 and its sgRNA. But in all cases, an efficient means to deliver the RCas9 PAMmer to the appropriate tissues must be identified. By limiting the expression or delivery of either Cas9 or the sgRNA to the tissue of interest and conducting systemic administration of the PAMmer, it may be possible to achieve tissue-specific RCas9 activity. Highly-stable modified oligonucleotides such as 2′-*O*-(2-methoxyethyl)-RNA have supported effective delivery and targeting of antisense RNAs in vivo [55–57] and may prove useful in the RCas9 system as well. Thus, while these and similar approaches have been used to deliver one or two components of the RCas9 system in vivo, it remains to be seen which combination allows effective reconstitution of all three components. Further modifications in the PAMmer will be required to prevent destruction of the target RNA due to recognition by RNAse H, the cellular enzyme that degrades RNA in RNA-DNA hybrids. Careful adjustment of the PAMmer length and modifications will be important to maintain targeting specificity while avoiding recruitment of the RNAi machinery. Although RCas9 does not appear to cleave DNA in vitro, it remains to be seen if inadvertent DNA targeting may occur in vivo. Ultimately, the success of RCas9 in vivo will ultimately rely on its specificity and whether RCas9 destabilizes the target RNA or interferes with its translation.

Recently in press

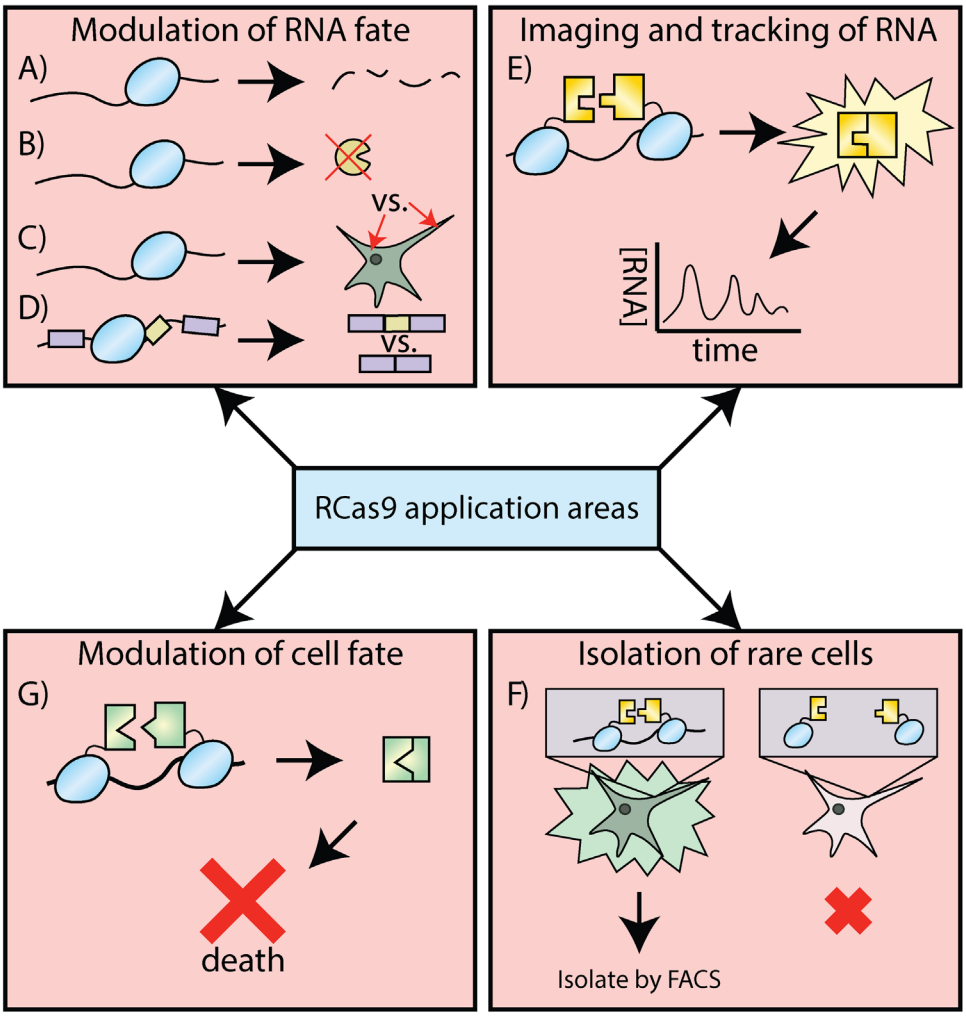


Figure 2. Summary of potential RCas9 application areas. **A–D:** describe means by which RNA fate can be manipulated by the RCas9 system. **A:** With a nuclease-active version of Cas9, siRNA-intractable RNA targets could be cleaved. **B:** Conversely, gene expression could be amplified by tethering factors that prevent degradation of target RNAs. **C:** By fusing Cas9 to a trafficking agent, RNAs could be forced to different sites of action in the cell for local translation or other activities. **D:** The processing of pre-mRNAs could be modulated by fusing Cas9 with a splicing factor to force differential exon choice. **E:** Along with altering RNA fate, RCas9 could be used to track RNA abundance in time with split luminescent or fluorescent proteins whose complementation is guided by binding of adjacent Cas9 proteins on RNA. **F:** Split fluorescent proteins could also be used to reveal rare cells by their RNA content for isolation by FACS and subsequent study. **G:** Finally, split toxic proteins or proteins that transform prodrugs to their active form could also be complemented in an RNA-dependent manner via fusion to Cas9.

useful in compartments or organelles where the RNAi machinery is not present or active. Further, the high affinity of RCas9 for RNA and dual recognition by both the sgRNA and PAMmer may allow more specific RNA depletion than siRNAs or antisense oligonucleotides. Table 1 compares this and other applications of RCas9 to current methods and Table 2 compares RCas9 for RNA knockdown to RNAi in greater detail.

Effective ways to enhance rather than decrease gene expression have been elusive. By fusing Cas9 to a protein factor that stabilizes mature messenger RNAs, it may be possible to enhance protein production from particular transcripts (Fig. 2B). Another permutation of the CRISPR-Cas system called CRISPR interference (CRISPRi) relies upon transcription modulators fused to a nuclease-null Cas9 (dCas9) [20, 21] that can enhance or repress gene expression by binding to particular genomic loci. While capable of strongly influencing gene expression, this approach does not allow isolation of the effects of RNA and protein gene products. By fusing Cas9 to translation enhancing factors, RCas9 may allow enhancement of protein expression of specific genes without altering RNA abundance in order to measure the specific importance of the protein gene product.

Another means by which cells control gene product activity is through the localization of RNA. In neurons, cell somata can be separated from synapses by centimeters or more, which presents a challenge to accumulating synaptic proteins at sufficient concentrations. After export from the nucleus, mRNAs involved in synaptic structure and activity such as postsynaptic density protein 95 (PSD-95) [22] are

Table 1. Summary of RCas9 potential applications

Application	State of the art	Limitations	RCas9-based approach	Main area of innovation
Targeted RNA knockdown (Fig. 1A)	siRNA, antisense oligonucleotides.	Efficiency limited by access to RNA silencing machinery and dependence on RNA structure.	Natural nucleolytic activity of Cas9.	Strong binding of Cas9 to target RNA may allow better knockdown efficiency; may allow knockdown in compartments lacking RNAi machinery.
RNA stabilization (Fig. 1B)	Coding region of GOI placed within stabilizing UTR contexts.	Requires targeted genetic manipulation or exogenous expression of GOI.	dCas9 fused to RNA stabilizing factor.	Potential first means to stabilize any unlabeled RNA.
RNA localization alteration (Fig. 1C)	*Cis*-acting sequence tags incorporated into transcript; these recruit tagged exogenous or endogenous localization factors.	Requires targeted genetic manipulation or exogenous expression of GOI.	dCas9 fused to RNA trafficking protein.	The high affinity of RCas9 for RNA could enable control of endogenous RNA localization.
RNA splicing alteration (Fig. 1D)	PUF proteins fused to splicing factors or splicing factor access blocked with antisense oligonucleotides.	PUFs limited to 8 base recognition sequences, oligonucleotides limited to splicing factor loss-of-function.	dCas9 fused to splicing factor targeted adjacent to or inside exons.	Potentially more specific alteration of splicing allowing either gain- or loss-of-function.
Imaging of RNA localization (Fig. 1E)	MS2 or Spinach labeling of RNA in conjunction with MS2-GFP protein or Spinach fluorophore.	Requires modification of target RNA.	dCas9 fused to fluorescent protein or split fluorescent protein.	May be effective means of revealing localization of any unlabeled RNA.
Time-resolved RNA measurements (Fig. 1E)	Incorporation of fluorescent or luminescent reporter at genomic locus near GOI.	Requires genetic modification.	dCas9 fused to split fluorescent or luminescent protein.	May be first means for time-resolved gene expression measurement without genetic modification.
Isolation of rare cells based on gene expression (Fig. 1F)	Identification of surface markers and antibodies for FACS.	Requires known surface marker for cell type of interest.	dCas9 fused to split fluorescent protein.	There are currently no high-sensitivity means to measure RNA content in live cells.
Death induction based in response to gene expression (Fig. 1G)	Incorporation of toxic protein at genomic locus near GOI.	Requires genetic modification, limited therapeutic potential.	dCas9 fused to split toxic protein.	Potentially first means to programmably target RNA profiles for death induction.
Substrate shuttling	Fusion of enzymes or incorporation of protein/protein interaction partners to create enzyme concatamers.	Results in constitutive substrate shuttling.	dCas9 fused to members of synthetic pathway targeting adjacent sites on an RNA.	First means to control substrate shuttling based upon RNA abundance.

Table 2. Comparison of RNAi and RCas9 for gene knockdown

	RNAi	RCas9
Specificity	Specificity determined by at most ~21 RNA nucleotides.	Target recognized by both 20 nucleotides within the sgRNA and the 20+ nucleotide PAMmer.
Components	Engages endogenous RNA-induced silencing complex (RISC); requires delivery of siRNA only.	Requires delivery of Cas9 protein, sgRNA, and PAMmer oligonucleotide.
Localization	RISC mainly cytoplasmic; targeting nuclear RNAs difficult.	RCas9 potentially active in both nucleus and cytoplasm.
Influence of RNA structure	Efficiency dependent on RNA accessibility and structure.	Cas9's helicase activity may allow recognition of structured RNA sequences.

transported through dendrites, where they are translated near their site of action (Fig. 2C). By fusing Cas9 to a transport factor, the RCas9 system could be used to force transport to a chosen region of the cell such as pre- or postsynaptic terminals. In the case of regeneration of neuronal processes after injury, there is some evidence that localization of RNAs that encode cytoskeletal components are critical to regrowth of axons [23, 23]. The ability to manipulate RNA localization in this context could be an important part of a regenerative therapy.

RCas9 could also be used to alter the composition of RNAs. Pre-mRNA splicing is a vital step in mRNA biogenesis and tethering of splicing factors to the pre-mRNA has been shown to alter the inclusion or exclusion of sequences [25]. For instance, the splicing factor RBFOX2 has been shown to influence inclusion of exons depending on whether it binds up- or downstream of alternative exons [26–28]. By carefully choosing splicing factors and fusing them to Cas9, it may be possible to create designer splicing factors whose influence on splice site choice can be determined by RCas9 sequence binding. For instance, spinal muscular atrophy is caused by deletion of the gene *SMN1* resulting in neuron death, but there is evidence that forced alteration of SMN2 splicing in SMN1-deficient cells can produce a SMN2 isoform that reconstitutes the activity of SMN1 [29]. This type of targeted splicing alteration could be used to reverse a variety of diseases caused by aberrant splicing [30].

These are just a few examples of RCas9's potential to modulate cellular RNA composition and cell behavior. As universal nucleic acid-recognitions proteins, Cas proteins could also allow targeting of particular RNAs to genomic loci. For instance, long non-coding RNAs (lncRNAs) can recognize particular genomic loci and guide associated chromatin-modifying factors to dramatic effect on genomic organization [31]. Considering this emerging principle, RCas9 fused to another Cas protein that utilizes an orthogonal sgRNA could be used to alter genome organization in a similar manner. By bringing targeted RNA in close proximity to a genomic locus of choice, this DNA- and RNA-binding Cas fusion could support studies of the function of lncRNAs in any genomic context. The use of multiple Cas proteins with orthogonal sgRNAs could also allow simultaneous and distinct alteration of multiple RNAs for instance by utilizing both nuclease-null and active Cas proteins. However, it is currently unclear whether other Cas proteins are capable of RNA recognition, so it remains to be seen if RCas9 can target multiple RNAs simultaneously [32].

Imaging applications

Several RNA recognition tools developed recently have enabled imaging of specific RNA species in live cells but suffer from several shortcomings (see [33] for an excellent review). In a manner analogous to visualizing proteins through fusion with fluorescent proteins, a set of RNA-based systems that rely on sequence tags incorporated in the RNA of interest can allow RNA visualization. These tags are specifically recognized by a protein moiety that binds strongly and specifically to the RNA tag. One popular approach utilizes bacteriophage MS2 coat protein (MCP) fused to a fluorescent protein [34] recognizing a short RNA structural motif (a so-called "hairpin"). The low signal-to-noise ratio due to background fluorescence from unbound probe can be improved by incorporating long arrays of tandemly repeated recognition elements, an approach that has allowed effective imaging of highly abundant RNAs in live cells [35], but there is concern that such large tags can significantly perturb typical RNA behavior. An alternative approach is the incorporation of artificial RNA sequence tags that are bound by an exogenous small molecule fluorophore [36, 37]. By immobilizing the fluorophore in this aptamer tag, fluorescence signal can be generated by increasing quantum yield, separating a fluorophore-quencher pair, or by FRET [36–39]. A third approach to suppress background fluorescence is the expression of two polypeptides that reconstitute a functional fluorescent protein when recruited to an RNA target by taking advantage of natural or artificial RNA-binding domains [40]. While all three methods have been tremendously useful and are widely used to study dynamics of RNA transport and localization, they require a tagged version of the RNA of interest introduced either through genetic modification of the endogenous locus or by forced expression of an exogenous tagged version of the RNA. To illustrate, cells derived from a knock-in mouse harboring 24 MS2 hairpins in the 3′-untranslated region (UTR) of the beta-actin gene allowed the real time visualization of transcription from the modified allele, including the observation of transcriptional bursting upon serum stimulation [41]. However, the MCP–GFP fusion proteins need to be delivered to cells exogenously and this system is limited to only highly expressed RNAs. Furthermore, incomplete occupation of the MS2 hairpins reduces local signal and generates significant background noise due to unbound probe [42]. Another technology, molecular beacons, allows imaging of unmodified transcripts but suffer from high noise and cumbersome delivery [43]. RCas9 may circumvent these issues by allowing direct recognition of untagged RNAs with high specificity and low noise.

We imagine RCas9 applications that allow visualization of the abundance and/or localization of one or more endogenous RNAs simultaneously. By fusing nuclease-null Cas9 to a fluorescent protein, it could be possible to visualize the localization of particular RNAs or RNA splice variants. Alternatively, a pair of Cas9 proteins could be fused to halves of split fluorescent protein such as Venus [44] and targeted to adjacent sites on an RNA (Fig. 2E). This might allow visualization of RNA localization with lower background than an intact fluorescent protein or measurement of the RNA content of individual cells. This split-protein approach could also be used to target adjacent exons in a differentially spliced transcript, allowing identification and isolation of individual cells that express particular RNA splice isoforms. The identification of Cas proteins with orthogonal sgRNAs could allow targeting of multiple transcripts for localization or abundance measurements simultaneously, allowing multiplexed, live-cell measurement of RNA dynamics of individual cells.

The applications of endogenous RNA localization and abundance measurements in live cells are numerous. For example, characterization of somatic stem cells remains difficult because few surface markers exist for cell

sorting-based identification and purification of these rare cells. Gene expression profiling remains the most effective way to identify rare cell types and RCas9 could enable this type of nondestructive measurement so that rare cells can be preserved, expanded, and studied in isolation.

RNA localization is important in cellular response to injury, stress, and some behaviors that promote cell polarity such as extension of neuronal processes. Stress granules are a type of RNA and protein cluster that sequester mRNA and protein and typically form in response to oxidative stress, heat, viral infection, or hypoxia [45]. Aberrant formation of RNA granules is implicated in many diseases, but the RNA components of these structures are only beginning to be described. We expect that a means to image endogenous RNA trafficking to stress granules would support investigation of stress granule roles in health and disease. In order to understand the importance of RNA granules in disease, RCas9 could allow time-resolved measurements of granule formation in response to stress, disease, or in drug screens where RNA localization may play a role in disease progression or regeneration of damaged tissues.

Synthetic biology applications

Synthetic biology is centered around the development of biological systems that have industrial, clinical, or other technological utility. Like all engineering-oriented disciplines, the development of modularized systems is a major focus so that flexible platforms can be tuned for diverse applications. The highly modular and programmable nature of the RCas9 system marks its potential as a platform technology in synthetic biology. For instance, it may be possible to fuse split enzymes to Cas9 proteins whose activity is reconstituted upon binding to a target RNA such as complementation of split death-inducing proteins after detection of a cancer-linked RNA (Fig. 2G). Further, it may be possible to re-engineer pathways involving successive protein/protein interactions by using RNA to scaffold interactions among Cas9 fusion proteins. A scaffold protein that binds kinases and their substrates has been shown to strongly influence the output of a signaling pathway [46], and RCas9 could allow scaffolding of protein-protein interactions to control signaling in a gene expression-dependent manner. Another group used tethering of enzymes involved in the production of the drug precursor mevalonate, thereby increasing production of this small molecule [47]. In principle, strong co-binding of Cas9 fusion proteins on a target RNA could lend a new level of control over successive protein interactions or shuttling of metabolites.

Conclusions, general concerns, and alternative approaches

Progress in RNA targeting methods from their beginnings, when RBP domains were adapted to serve as sequence specificity determinants, to RCas9, with its target recognition by simple nucleic acid hybridization, is poised to closely parallel the development of DNA targeting technology. Here, zinc finger and TAL effector nucleases have recently given way to DNA recognition by the Cas9-bound sgRNA. While for DNA targeting applications, Cas9 and its sgRNA are sufficiently stable and nontoxic in mammalian cells, it remains to be seen whether all three components of the RCas9 system (Cas9, sgRNA, and PAMmer) can be delivered efficiently and, if so, successfully cooperate to bind target RNA. Alternative approaches to RNA-programmed RNA recognition are on the horizon. Type III-B CRISPR-Cas systems are known to target and cleave RNA as part of their normal activities in bacterial immunity. The effector complexes of these CRISPR systems from *Thermus thermophilus* [48] and *Pyrococcus furiosus* [49] have been characterized in detail and their nucleolytic activities reconstituted in vitro. While the natural ability of these complexes to recognize RNA is appealing, each complex is composed of 1–4 copies of six different proteins, which could pose challenges for its reconstitution in vivo. Cas9 from *Francisella novicida* targets a particular endogenous RNA in this organism in a RNA-guided manner, although the flexibility of this system to target chosen RNAs remains unclear [50].

The ability of RCas9 to recognize untagged, endogenous RNA via simple base-pairing represents a major advance in RNA targeting and is particularly critical in diagnostic or therapeutic applications. If the system can be delivered efficiently to cells, cooperate to recognize RNA, and avoid unwanted destabilization or alteration of target RNA while also avoiding targeting of genomic DNA and off-target transcripts, one can imagine many new applications of RCas9 in basic and applied biology and in medicine.

Acknowledgements

This review was partially supported by NIH grants NS075449, HG004659, HG007005 to GWY.

The authors have declared no conflict of interest.

References

1. The ENCODE Project Consortium 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
2. **Pasca SP, Portmann T, Voineagu I, Yazawa M**. 2011. Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat Med* **17**: 1657–62.
3. **Lu J, Getz G, Miska EA, Alvarez-Saavedra E**, et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435**: 834–8.
4. **MacKenzie TA, Schwartz GN, Calderone HM, Graveel CR**, et al. 2014. Stromal expression of miR-21 identifies high-risk group in triple-negative breast cancer. *Am J Pathol* **184**: 3217–25.
5. **Gerstberger S, Hafner M, Ascano M, Tuschl T**. 2014. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv Exp Med Biol* **825**: 1–55.
6. **DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL**, et al. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**: 245–56.
7. **Renton AE, Majounie E, Waite A, Simon-Sanchez J**, et al. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**: 257–68.
8. **O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A**, et al. 2014. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**: 263–6.
9. **Beuth B, Pennell S, Arnvig KB, Martin SR**, et al. 2005. Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J* **24**: 3576–87.

Recently in press

10. **Braddock DT, Louis JM, Baber JL, Levens D**, et al. 2002. Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415**: 1051–6.
11. **Laird-Offringa IA, Belasco JG**. 1995. Analysis of RNA-binding proteins by in vitro genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target. *Proc Natl Acad Sci USA* **92**: 11859–63.
12. **Wang X, Zamore PD, Hall TM**. 2001. Crystal structure of a Pumilio homology domain. *Mol Cell* **7**: 855–65.
13. **Filipovska A, Razif MF, Nygard KK, Rackham O**. 2011. A universal code for RNA recognition by PUF proteins. *Nat Chem Biol* **7**: 425–7.
14. **Wang Y, Cheong CG, Hall TM, Wang Z**. 2009. Engineering splicing factors with designed specificities. *Nat Methods* **6**: 825–30.
15. **Zhang W, Wang Y, Dong S, Choudhury R**, et al. 2014. Treatment of type 1 myotonic dystrophy by engineering site-specific RNA endonucleases that target (CUG)(n) repeats. *Mol Ther* **22**: 312–20.
16. **Mali P, Yang LH, Esvelt KM, Aach J**, et al. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823–6.
17. **Cong L, Ran FA, Cox D, Lin S**, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–23.
18. **Cencic R, Miura H, Malina A, Robert F**, et al. 2014. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9**: e109213.
19. **Kuscu C, Arslan S, Singh R, Thorpe J**, et al. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**: 677–83.
20. **Gilbert LA, Larson MH, Morsut L, Liu Z**, et al. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**: 442–51.
21. **Qi LS, Larson MH, Gilbert LA, Doudna JA**, et al. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173–83.
22. **Muddashetty RS, Nalavadi VC, Gross C, Yao X**, et al. 2011. Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. *Mol Cell* **42**: 673–88.
23. **Shestakova EA, Singer RH, Condeelis J**. 2001. The physiological significance of beta -actin mRNA localization in determining cell polarity and directional motility. *Proc Natl Acad Sci USA* **98**: 7045–50.
24. **Donnelly CJ, Willis DE, Xu M, Tep C**, et al. 2011. Limited availability of ZBP1 restricts axonal mRNA localization and nerve regeneration capacity. *EMBO J* **30**: 4665–77.
25. **Graveley BR, Maniatis T**. 1998. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell* **1**: 765–71.
26. **Lovci MT, Ghanem D, Marr H, Arnold J**, et al. 2013. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol* **20**: 1434–42.
27. **Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S**, et al. 2014. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* **6**: 1139–52.
28. **Yeo GW, Coufal NG, Liang TY, Peng GE**, et al. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**: 130–7.
29. **Hua Y, Vickers TA, Okunola HL, Bennett CF**, et al. 2008. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am J Hum Genet* **82**: 834–48.
30. **Nissim-Rafinia M, Kerem B**. 2002. Splicing regulation as a potential genetic modifier. *Trends Genet* **18**: 123–7.
31. **Geisler S, Coller J**. 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* **14**: 699–712.
32. **Esvelt KM, Mali P, Braff JL, Moosburner M**, et al. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10**: 1116–21.
33. **Rath AK, Rentmeister A**. 2014. Genetically encoded tools for RNA imaging in living cells. *Curr Opin Biotechnol* **31C**: 42–9.
34. **Bertrand E, Chartrand P, Schaefer M, Shenoy SM**, et al. 1998. Localization of ASH1 mRNA particles in living yeast. *Mol Cell* **2**: 437–45.
35. **Park HY, Lim H, Yoon YJ, Follenzi A**, et al. 2014. Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* **343**: 422–4.
36. **Paige JS, Wu KY, Jaffrey SR**. 2011. RNA mimics of green fluorescent protein. *Science* **333**: 642–6.
37. **Strack RL, Disney MD, Jaffrey SR**. 2013. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat Methods* **10**: 1219–24.
38. **Sunbul M, Jaschke A**. 2013. Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. *Angew Chem Int Ed Engl* **52**: 13401–4.
39. **Shin I, Ray J, Gupta V, Ilgu M**, et al. 2014. Live-cell imaging of Pol II promoter activity to monitor gene expression with RNA IMAGEtag reporters. *Nucleic Acids Res* **42**: e90.
40. **Rackham O, Brown CM**. 2004. Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J* **23**: 3346–55.
41. **Lionnet T, Czaplinski K, Darzacq X, Shav-Tal Y**, et al. 2011. A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* **8**: 165–70.
42. **Fusco D, Accornero N, Lavoie B, Shenoy SM**, et al. 2003. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr Biol* **13**: 161–7.
43. **Tyagi S, Kramer FR**. 1996. Molecular beacons: Probes that fluoresce upon hybridization. *Nat Biotechnol* **14**: 303–8.
44. **Ozawa T, Natori Y, Sato M, Umezawa Y**. 2007. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* **4**: 413–9.
45. **Kedersha N, Anderson P**. 2007. Mammalian stress granules and processing bodies. *Methods Enzymol* **431**: 61–81.
46. **Bashor CJ, Helman NC, Yan S, Lim WA**. 2008. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319**: 1539–43.
47. **Dueber JE, Wu GC, Malmirchegini GR, Moon TS**, et al. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27**: 753–9.
48. **Staals RH, Zhu Y, Taylor DW, Kornfeld JE**, et al. 2014. RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell* **56**: 518–30.
49. **Hale CR, Zhao P, Olson S, Duff MO**, et al. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945–56.
50. **Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL**, et al. 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**: 254–7.
51. **Swiech L, Heidenreich M, Banerjee A, Habib N**, et al. 2015. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat Biotechnol* **33**: 102–6.
52. **Maddalo D, Manchado E, Concepcion CP, Bonetti C**, et al. 2014. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**: 423–7.
53. **Dow LE, Fisher J, O'Rourke KP, Muley A**, et al. 2015. Inducible in vivo genome editing with CRISPR-Cas9. *Nat Biotechnol* in press DOI: 10.1038/nbt.3155
54. **Zuris JA, Thompson DB, Shu Y, Guilinger JP**, et al. 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**: 73–80.
55. **Meng L, Ward AJ, Chun S, Bennett CF**, et al. 2015. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518**: 409–12.
56. **Hua Y, Sahashi K, Hung G, Rigo F**, et al. 2010. Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev* **24**: 1634–44.
57. **Passini MA, Bu J, Richards AM, Kinnecom C**, et al. 2011. Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. *Science Transl Med* **3**: -72ra18.