Next stop for the CRISPR revolution: RNA-guided epigenetic regulators

Suhani Vora1,2,3, Marcelle Tuttle1, Jenny Cheng1 and George Church1,2

1 Wyss Institute for Biologically Inspired Design, Center for Life Sciences Boston, Boston, MA, USA
2 Department of Genetics, Harvard Medical School, Boston, MA, USA
3 Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Keywords
activation; CRISPR-associated protein 9; clustered regularly interspaced short palindromic repeats; epigenetic; repression; transcription

Introduction: CRISPR-associated protein 9, the catalyst

Epigenetics mediate which portions of a genome are open or closed for expression. Many fundamental biological phenomena are regulated at the transcriptional and epigenetic, rather than genetic, level. For example, epigenetics regulate organismal development, allowing a small number of isogenic embryonic stem cells to divide, differentiate, and specialize into the various tissues that constitute the human body. Tools to directly control gene expression and manipulate epigenetic marks will enable forward engineering of these processes. The ability to manipulate transcriptional and epigenetic state on demand will open doors to directing differentiation of pluripotent stem cells into specialized tissue types, to establishing causative relationships between expression state and phenotype, and to correcting disease phenotypes at the epigenetic level.

Numerous zinc fingers (ZFs) and transcriptional activator-like effectors (TALEs) have been fused to effector domains in order to programmably target and activate, Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins offer a breakthrough platform for cheap, programmable, and effective sequence-specific DNA targeting. The CRISPR-Cas system is naturally equipped for targeted DNA cutting through its native nuclease activity. As such, groups researching a broad spectrum of biological organisms have quickly adopted the technology with groundbreaking applications to genomic sequence editing in over 20 different species. However, the biological code of life is not only encoded in genetics but also in epigenetics as well. While genetic sequence editing is a powerful ability, we must also be able to edit and regulate transcriptional and epigenetic code. Taking inspiration from work on earlier sequence-specific targeting technologies such as zinc fingers (ZFs) and transcription activator-like effectors (TALEs), researchers quickly expanded the CRISPR-Cas toolbox to include transcriptional activation, repression, and epigenetic modification. In this review, we highlight advances that extend the CRISPR-Cas toolkit for transcriptional and epigenetic regulation, as well as best practice guidelines for these tools, and a perspective on future applications.

Abbreviations
AAV, adeno-associated virus; Cas9, CRISPR-associated protein 9; CRISPRa, CRISPR activation; CRISPR, clustered regularly interspaced short palindromic repeats; CRISPRi, CRISPR interference; dCas9, dead Cas9; DMD, Duchenne muscular dystrophy; KRAB, Kruppel-associated box; lncRNA, long noncoding RNA; SAM, synergistic activation mediator; sgRNA, single guide RNA; SpCas9, Streptococcus pyogenes Cas9; SunCas9, SunTag Cas9; SunTag, SUperNova Tag; TALE, transcription activator-like effector; ZF, zinc finger.
CRISPR-Cas9 guided epigenetic regulators

S. Vora et al.

repress, or epigenetically modify endogenous genes, laying the foundation for precision regulation of epigenetic state [1–8]. Unfortunately, inefficient production pipelines prevent widespread use of these tools.

Zinc fingers engineered to target an arbitrary sequence must be constructed in large oligonucleotide pool arrays that are expensive, labor-intensive to generate, and are not guaranteed to be specific to the intended sequence [9,10]. Similarly TALEs, while a powerful platform, pose a significant challenge for assembly, sequence verification, and genetic stability due to their highly repetitive nature. Modern methods have managed to automate labor-intensive steps of construction [11]. Yet, even with an automated approach, the construction of a TALE library designed to target every protein coding locus in the human genome would require over 1 million dollars and 200 days [11].

Recently discovered, clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems have provided a quantum leap forward for programmable DNA-binding tools. In nature, CRISPR-Cas systems serve as archaeal and bacterial immune systems [12–17]. Although five different types of CRISPR-Cas systems have been discovered to date, the type II system has almost exclusively been used due to the small number of components required for DNA targeting [18]. Programmable DNA targeting by the type II CRISPR-Cas system only requires three components: CRISPR-associated protein 9 (Cas9) nuclease, CRISPR RNA (crRNA), and transactivating crRNA (tracrRNA) [19–22].

To initiate programmed DNA targeting, Cas9 binds both crRNA and tracrRNA in an RNA-duplex. The first 20 nucleotides of the crRNA guide Cas9 to cleave any complementary genomic sequence located next to a short protospacer adjacent motif (PAM) [23,24]. A convenient fusion of the crRNA and tracrRNA produces a chimeric single guide RNA (sgRNA) that is sufficient to target Cas9 as well [19].

Sequence-specific sgRNA are quick and easy to construct [25], and scaling up to produce tens of thousands of sgRNA is a straightforward process. Today, a library of sgRNA targeting every protein coding locus in the human genome can be constructed via chip-based oligonucleotide synthesis for roughly 5000 dollars, within a few weeks [25–27]. As such, the two component Cas9 and sgRNA system has been adopted as the standard genome editing tool for a wide range of organisms [19,28,29].

Following the widespread adoption of Cas9 as a programmable tool for genome editing, several groups including our own mutated residues involved in DNA cleavage located within each of two nuclease domains of Cas9 [19,30]. Mutation of these catalytic residues generated a nuclease-null ‘dead’ Cas9 (dCas9) variant, disabling its ability to cleave DNA, while retaining the ability to bind DNA in a sequence-specific manner. Fusion of the dCas9 protein to effector domains, such as transcription enhancing factors, repressors, and epigenetic modulators, has enabled unprecedented ease of eukaryotic transcriptome and epigenome manipulation. Here, we cover the evolution of Cas9 activators, repressors, and epigenetic regulators, as well as the challenges and opportunities associated with their use.

Repurposing Cas9 as a genetic Swiss Army knife

Soon after initial studies demonstrated the simplicity and effectiveness of Cas9-based genome editing in human cell lines [20–22], several groups extended the Cas9 platform of RNA-guided genome editing to include activation, repression, and epigenetic modification—creating a genetic Swiss Army knife.

Evolution of robust second-generation Cas9-based activators

First-generation activators

To enable eukaryotic genome regulation, previously established transcriptional activators have been fused to a dCas9 ortholog from the bacteria Streptococcus pyogenes (dSpCas9). Initial studies reported dSpCas9 fusions to the commonly used VP64 activator (four tandem repeats of the herpes simplex virus protein 16, VP16) to create a first-generation dSpCas9-VP64 activator (Fig. 1A). Unfortunately, when targeted to the promoters of endogenous genes to induce gene expression, the level of gene induction was quite modest and lower than that of equivalent TALE-based activators [5,30–33]. A subset of these studies [30–32] indicated the clear necessity for multiple guides in order to up-regulate genes of interest, demonstrating additive and synergistic effects of activating genes with up to 10 guides.

However, many applications such as viral delivery of tools in vivo, and library-based screening methods, benefit significantly from the compactness and simplicity of using a single guide per target. Therefore, it was imperative to build stronger, more potent activators to activate genes robustly using only a single guide.

Enhanced activators

The key to building improved activators lay in the observation that multiple weak activator domains,
concentrated at a single locus, greatly enhance the level of gene expression achieved. Taking advantage of this observation, many groups fused a large number of VP16 protein repeats directly to dCas9. Multiple enhanced activator tools emerged, including dSpCas9-VLP160 (10 repeats of VP16) [34], dSpCas9-VP192 (12 repeats of VP16) [35], and VP64-dSpCas9-BFP-VP64 (four repeats of VP16 on the N terminus of dSpCas9, and blue fluorescent protein plus four more repeats of VP16 on the C terminus) [36] (Fig. 1B).

Each technology created a high local concentration of activation domains by directly chaining them to dCas9, as opposed to relying on multiple guides to bring separate weaker activators together. Through this strategy, these groups successfully reduced the average number of guides required to induce potent gene expression.

Second-generation activators

Second-generation activator tools explored several novel methods of enhancing potency. Three parallel methods included: recruiting activation factors in trans to the Cas9 scaffold to avoid expressing long stretches of repetitive proteins, insertion of protein-binding RNA-aptamer hairpins into the sgRNA to facilitate activator recruitment in trans, and use of heterogeneous activator domains to more effectively recruit transcriptional machinery.
Biological signals are often amplified by recruitment of multiple effectors proteins to the site of action. The SUperNova tagging (SunTag) system employs 10 repeats of a short epitope tag fused directly to dCas9 (SunCas9), allowing in trans expression and recruitment of 10 single-chain variable fragments (scFvs) fused to VP64 [37] (Fig. 1C). As a result, single guides are able to induce robust gene expression, activating genes up to 45× with the SunCas9 system. It is important to note that ‘fold-activation’ metrics are imperfect to determine the true potency of a tool, as the ability to activate a gene depends on factors including baseline level of gene expression and cell line [38]. Therefore, a genuine metric of potency is relative efficacy compared to the gold standard activator, dSpCas9-VP64. Demonstrating the relative efficacy of SunCas9, the tool showed 3× and 45× improvement over dSpCas9-VP64 when targeted to a pair of endogenous genes, respectively. A parallel study uncovered fundamental rules about SunCas9 sgRNA placement, with most effective guides for activation being optimally located between −400 and −50 base pairs (bp) upstream of the transcriptional start site (TSS) [39].

Alternatively, RNA-aptamer-based strategies have also been used to recruit activators onto the sgRNA [30,40,41]. In the earliest example of an sgRNA-based recruitment strategy, RNA hairpins from the male-specific bacteriophage-2 (MS2) were repetitively fused to the 3’ end of an sgRNA. In tandem, an MS2 coat protein (MCP), which binds each MS2 hairpin as a homodimer, was fused to the VP64 activator domain. When the modified guide, MCP-VP64 fusion protein, and dSpCas9 are coexpressed, Cas9 binds the decorated sgRNA which in turn recruits MCP-VP64 to activate expression of the targeted gene [30].

Progressing the sgRNA-based recruitment strategy, the synergistic activation mediator (SAM) system incorporates two MS2 hairpins into exposed loops within the sgRNA and fuses MCP to a novel chimeric activator p65-HSF1 (p65 is a subunit of the ubiquitous NF-kB transcription factor complex, and HSF1 is heat shock factor 1 responsible for transcribing genes in response to temperature stress). The chimeric MCP-p65-HSF1 activators bind each MS2 hairpin as a set of homodimers, resulting in four copies of MCP-p65-HSF1 scaffolding onto the SAM sgRNA, bound to dSpCas9-VP64 (Fig. 1D) [40]. With this advanced recruitment strategy, single guides were shown to successfully induce high levels of expression, with a median gain of 105× improvement above the gold standard dSpCas9-VP64 activator. Similar to the SunCas9 system, an optimal window for targeting the dSpCas9 SAM activator was found to be between −200 and +1 bp relative to the TSS.

In a simplified yet robust system, three separate activation domains were fused in tandem directly to dSpCas9. In nature, transcriptional activation occurs through multiple transcriptional activating factors binding to the promoter and enhancers surrounding a gene, which in turn signal the mediator complex to facilitate RNA polymerase II transcription initiation and elongation. By employing a set of three diverse transactivation domains, VP64, p65, and Rta, each predicted to initiate independent contacts with mediator, the dSpCas9-VP64-p65-Rta (dSpCas9-VPR) system was shown to work 20× to 300× better than the gold standard dSpCas9-VP64 tool with four guides [38] (Fig. 1E).

**Epigenetic activator**

In addition to transcription-based activation of gene expression, a tool for Cas9-mediated epigenome editing to establish an activating mark has been created as well. The catalytic histone acetyltransferase (HAT) core domain of the human E1A-associated protein p300 was fused to Cas9 to catalyze the addition of a H3K27 acetylation mark, allowing for activation of gene expression when targeted to promoters, proximal enhancers, and distal enhancers (Fig. 1F) [42]. The dSpCas9-p300 core fusion was demonstrated to be up to 6× more potent than the gold standard activator dSpCas9-VP64 when targeted to promoter regions. In addition, dSpCas9-p300 induced 5× to 270× increased gene expression of downstream genes in the locus control region (LCR) of the human B-globin locus, by directly acetylating distal enhancer regions. When bound to these same regions, dSpCas9-VP64 was unable to influence transcription, highlighting the distinct mechanism of activation orchestrated by the epigenetic regulator.

**Development of first-generation Cas9 repressors**

**Introduction of CRISPR interference**

An additional utility of Cas9 was also displayed through repression of gene expression in *Escherichia coli* and human cells. The system, consisting of the dCas9 protein and an sgRNA, was coined CRISPR interference (CRISPRi). CRISPRi can silence both transcription initiation as well as elongation, by targeting either the −35 box of the promoter (in bacteria only) or downstream regions of the gene itself (Fig. 2A) [43]. Several key insights were
made in this seminal study: the location of the sgRNA along the promoter or gene was critical for efficiency and could control transcriptional output, the use of multiple sgRNA allowed for combinatorial enhancement of the regulatory effect, and multiplexed guides could be used to easily repress multiple genes simultaneously. While the repressive effect in this seminal study was quite dramatic for bacterial gene repression, only a modest 1.9× to 2.7× down-regulation of expression was described when targeting genes in human cells.

**Stronger epigenetic repressors**

A subsequent study attempted to enhance the repressive effect of CRISPRi in human cells by fusing the Kruppel-associated box (KRAB) domain of Kox1 to dCas9, in order to recruit chromatin modifying complexes and more efficiently silence transcription [33] (Fig. 2B). The fusion showed 15-fold repression of an integrated GFP reporter in HEK293T cells, with the dSpCas9-KRAB integrated as well.

Genomic integration of both repressor and guide facilitated long-term silencing, with no detectable GFP expression left in 95% of the population after 14 days. However, transient expression of the repressor was not shown, leaving a gap in our understanding of the ‘epigenetic’ capacity of the repressor. While the term ‘epigenetic’ traditionally often implies heritability, stable integration of the repressor only confirms the ability to regulate transcription while the tool is actively expressed in the target cell line. Transient expression data would clarify whether the repressive epigenetic marks made by dCas9-KRAB are self-perpetuating, and therefore capable of maintaining long-term transcriptional silencing after the repressor has been removed. The study did show a strong correlation between dCas9 and dCas9-KRAB repression, indicating that strength of this binding event might be a limiting factor for all CRISPRi-mediated repression.

In another approach, four copies of the mSin3 domain (termed SID4X) were fused in tandem to the dCas9 scaffold, and shown to repress the endogenous gene SOX2 in HEK293T cells more than 2×, while dCas9 showed no effect [4]. More endogenous targets, however, should be tested with this repressor to determine its relative efficacy, and its efficacy versus the benchmark dSpCas9-KRAB fusion. However, stronger repressors than the dSpCas9-KRAB fusion are likely to be found. Strategies used to upgrade activator tools clearly apply to repressors as well; perhaps, a dCas9 fusion to a series of KRAB domains or a combination of repressive domains would allow for even stronger repression.

While the dSpCas9-KRAB tool relies on local heterochromatinization [44] to silence expression, alternative domains silence gene-specific enhancers by removing specific activating epigenetic marks. Histone demethylase LSD1, fused to a small dCas9 ortholog from *Neisseria meningitis* (dNmCas9), removes activating H3K4 methylation marks leading to enhancer decommissioning and gene repression [45] (Fig. 2C).

Working specifically through enhancer inactivation, as opposed to general heterochromatinization, the tool allows for a new level of epigenetic precision to control transcriptional state.

**Perspective on the emergence of Cas9 epigenetic modifiers**

Despite the advances made in Cas9-based histone modifiers, no common DNA-modifying demethylase or methyltransferase effectors have been generated with Cas9. Such fusions work when fused to TALEs as shown with TALE-Tet1 dioxygenase (providing a key step to demethylation) and TALE-DNMTs (DNA methyltransferases) [7,8]. The development of these tools is a crucial step for expanding our Cas9 regulatory toolkit.

**Newly minted multi-mode Cas9 regulators**

Recently created tools have allowed for simultaneous activation and repression of multiple target genes, using a single Cas9 protein [41,46,47]. Two studies described a method for truncated sgRNA-based concurrent activation, repression, and cutting [46,47]. Previously, it was established that mismatch-containing guides allowed for binding, but prevented cutting [48]. This is due to conformational changes in Cas9. When there are roughly six or more mismatches on the 5’ end of the sgRNA, Cas9 may still bind to the target site, but will not rotate into an active conformation to cut the target [49].

Similarly, sgRNA with truncated targeting sequences of 14nt instead of 20nt disable the cutting ability of Cas9 as well, while still allowing for wild-type Cas9 to bind. As a result, one can achieve activation, repression, and mutation for multiple target genes by deploying sgRNA of different lengths for each target site. For example, one can deploy a wild-type Cas9-activator fusion and a 14-nt truncated guide targeted to the promoter of a gene to activate expression without cutting. Simultaneously, one can target the fusion downstream of a TSS with a truncated sgRNA, enabling the fusion to act as a roadblock for RNA polymerase II read through without cutting. Finally, one can target the fusion to any site with a full-
length sgRNA, allowing for cutting. This was shown to work with both a nuclease-positive SpCas9-VPR and a stripped down version of the SAM system employing SpCas9 without VP64. A promising use of these tools will be the extension of pre-existing wild-type Cas9 mouse models [50,51] and difficult to generate Cas9 cell lines with the capacity to actuate transcriptional activation and repression as well.

**Major concerns**

With a flood of tools for genomic perturbation becoming available for the scientific community, several concerns have emerged. Major concerns include the relative potencies of each tool in comparison to one another or within various chromatin contexts, potential off-target effects caused by nonspecific Cas9 binding, and the ability to deliver the genes encoding these tools through the commonly used viral vectors for *in vivo* research and applications. Here, we address these concerns and offer suggestions as to how these issues can be overcome.

**Strength**

While regulatory tools have evolved in potency and sophistication, layers of contextual factors shroud the true efficacy of these tools. Studies detailing the development of these tools use a wide array of test bed cell lines, genes, sgRNA, and delivery methods that cannot be compared. In order to efficiently apply these tools in their best capacity, we must understand the cellular, genomic, and positional contexts in which each of the activators and repressors performs best. For example, two recent studies have indicated a negative effect of packed chromatin on the ability of Cas9 to bind specific genomic loci [52,53]. Additional studies exploring the effects of chromatin packing on second-generation Cas9 activators as well as epigenetic repressors would be useful to understand the optimal contexts in which to apply these effectors, and how best to engineer tools to overcome such barriers (such as the use of histone modifiers in combination with traditional transcriptional activators). Furthermore, a systematic comparison of available tools would be particularly beneficial at this time, as there is much redundancy, leading to ambiguity as to which tool should be adopted given a particular application or context [54].

**Specificity**

Although off-target mutations can occur throughout the entire genome [30,55], off-target transcriptional
activation or repression can only occur at a very small fraction of the genome [39,53]. The off-target guide would have to serendipitously target Cas9 to land within a 1000-bp region surrounding the TSS of a gene to induce or repress transcription. However, many researchers use previously published off-target guide-scoring algorithms to design ‘specific guides’ when applying Cas9 for transcriptional perturbations [55–58]. The most specific guide is unlikely to be the most potent [57]. Small targeting windows for transcription regulating tools do not afford a large number of guides from which to cull both a potent and ‘specific’ choice. Therefore, rejecting guides in favor of enhancing ‘specificity’ may ultimately be detrimental to the use of these tools unless it is clear that the rejected guides truly cause off-target transcriptional effects. It would be useful if guide-scoring algorithms were modified to differentiate between cutting- and binding-related activities when determining whether a guide will lead to an off-target effect.

**Delivery**

For delivery to somatic tissues in adult animal models, the adeno-associated virus (AAV) is a very attractive delivery vehicle. AAV vectors have relatively low immunogenicity, minimal risk of integration into the host genome, and a wide spectrum of tissue-specific serotypes [59–61]. Unfortunately, the payload limit for AAV is around 4.5 kb, while Sp Cas9 is 4.2 kb, leaving just enough room for small regulatory elements and necessitating use of an additional vector to deliver the sgRNA expression cassette [62]. Ideally, second-generation activators, repressors, and epigenetic effectors will be delivered for in vivo perturbation of gene expression as well. This task will prove more difficult than delivering Cas9 alone, due to the additional space required for regulatory effectors. It is imperative that smaller orthologs such as the 3.2 kb SaCas9 and St1Cas9 orthologs are included when building transcription and epigenetic effector tools to enable adaptation for in vivo applications [51,63]. Ultimately, either split systems must be engineered (with regulator parts split among multiple AAVs, and codelivered for in vivo assembly) or minimalist solutions will need to be identified to develop compact single vector systems.

**Current and future applications**

Speaking to the ease of use and utility of Cas9-based regulators, Cas9 transcriptional regulators have been used to tackle several previously challenging biological applications. Regulators have been used to deconstruct genetic networks that confer resistance phenotypes,mediate cellular differentiation by activating endogenous genes, activate latent HIV genomes to improve efficacy of antiretroviral therapy, and induce a morphological phenotype in vivo. Looking forward, we foresee applications to genome-wide screening of poorly understood classes of genes, as well as transcriptional and epigenetic paradigms for human gene therapy.

**Screening for resistance phenotypes**

A rich variety of epigenetic regulatory tools built around Cas9, combined with the economical production of short oligonucleotides to synthesize tens of thousands of sgRNA, have advanced the rate and depth with which genome-scale functional screens can be conducted. While genome-scale screens require manual labor and time, the amount of data and the potential to answer large biological questions are enticing to researchers. Even though screens using previous technologies, such as RNAi-based loss of function and cDNA-based gain of function screens, have been conducted extensively, Cas9 tools offer an undeniable new level of flexibility, multiplexing capability, and precision, while also providing the ability to target endogenous loci and regulate noncoding portions of the human genome.

To date, only three genome-scale screens have been published with Cas9 activators and repressors, two of which were conducted as a part of the same study [39,40]. The initial two screens were conducted with the dCas9-KRAB repressor, and the dCas9-SunTag activator. The screens included a loss of function (CRISPRi) study for genes that confer sensitivity to cholera–diphtheria fusion toxin, and a gain of function screen [CRISPR activation (CRISPRa)] for genes that confer resistance to the same toxin. These screens demonstrate a key benefit of Cas9-based screening. In order to truly understand genetic networks, it is necessary to learn the effects of both up-regulation and down-regulation on each specific gene. Cas9 transcriptional technology allows us to easily probe both functions and therefore understand gene regulatory networks at a deeper level. Another insight from these screens is the need to construct separate libraries for CRISPRa and CRISPRi screens due to the different gRNA positions needed for maximal effect. While it might be tempting to use the same library for a screen and save the effort of constructing and validating an additional library, there is a strong risk of losing potency and therefore potential hits.

**Cellular differentiation**

Early studies demonstrated the potential for dSpCas9-VP64 activators to activate developmentally relevant
genes such as *SOX17* for directed differentiation to endodermal lineage, and *OCT4* for direct reprogramming of somatic cells to a pluripotent state [64,65]. Unfortunately, the first-generation dSpCas9-VP64 tool lacked the potency required to induce high levels of regulatory factor expression and trigger cellular reprogramming [36,38,65]. As a result, enhanced and second-generation activators were engineered, enabling the first examples of CRISPR-Cas9 mediated cellular reprogramming and differentiation. When targeted to *Myod1* in mouse fibroblasts, the enhanced VP64-dSpCas9-BFP-VP64 activator directed cellular reprogramming to myocytes [36]. Additionally, our group employed the second-generation dSpCas9-VPR activator to activate master neuronal lineage regulators *NEUROD1* and *NEUROG2* in induced pluripotent stem cells (iPSCs,) and differentiate them into neuronal cells.

While improved activators allow full-fledged cellular reprogramming through the induction of master transcription factors, in reality most reprogramming protocols require the expression of multiple genes. Multiplexed gene activation would be ideal for reprogramming purposes. Yet, many genes are recalcitrant to Cas9 activator-based induction, making the process of applying tools very hit or miss. As previously mentioned under major concerns for these tools, there is a clear need to understand why activators work well in certain genomic contexts as opposed to others, and build tools that can be universally applied for transcriptional regulation.

**Human therapeutics**

Much of the excitement surrounding Cas9 is related to its predicted role in rapidly and efficiently developing genetic therapies. However, unlike Cas9-based strategies that rely on nuclease activity, Cas9 transcriptional regulators have the added advantage that there is no significant risk of introducing mutations to the host genome, while providing potentially reversible, short-term, or long-term epigenetic modifications.

Despite the vast potential for therapeutic applications of Cas9 regulators in therapeutic applications, very few studies on the topic have been published as of yet. An example of a potential therapeutic application includes use as an adjuvant for antiretroviral therapy to potentially cure HIV [66,67]. Current antiretroviral regimens prevent the spread of viral infection by targeting enzymes required for HIV to complete its life cycle of entry, integration, assembly, and budding. However, a fraction of integrated virus remains transcriptionally silent in the host genome, evading targeting by both antiretroviral drugs as well as the immune system. Attempts at dispensing chemical epigenetic factor inhibitors to prevent silencing have potentially widespread off-target effects on the remainder of the genome and have generally proven ineffective in clinical trials. In an alternative strategy, programmable Cas9 activators have been used to specifically induce expression of latent HIV genome [66,67]. Activation of silent HIV genomes might allow for a long sought after curative therapy in which cells harboring latent reservoirs of viral genomes are forced to express the viral genes they harbor, making them susceptible to antiretroviral drugs and allowing clearance by the immune system.

Furthermore, the use of Cas9-based transcriptional effectors allows for rapid screening of regulatory elements that can be targeted for highly specific activation of latent HIV-1 viral genomes, and thus provides a promising approach for a molecular biology-based antiretroviral strategy [66,67].

Beyond this application, not much has been done in the way of Cas9 regulator-based therapeutic applications. Nevertheless, several applications, such as the activation of therapeutic genes, anticancer genes, haploinsufficient genes, or repression of oncogenes, cancer drug resistance-conferring genes, and dominant negative genes all stand to benefit from the application of these tools. All can be immediately targeted without the need to screen for small molecule drugs, and without the concern of widespread off-target effects.

**In vivo applications**

As of yet, only one application of Cas9-based regulation has been conducted to induce a phenotype *in vivo*. The dSpCas9-VPR activator was applied to Wingless gene activation in *Drosophila*, altering the structure of wing pouches developed in larvae [68]. There are promising new avenues for investigation on the horizon. A Cas9 mouse has recently been created [50]; given the ability of the modified SAM system to actuate genome editing, activation, and repression with wild-type Cas9, it is likely that this mouse will be used to conduct transcriptional regulating studies as well [46,50]. Additional applications *in vivo* are expected soon, as this is too valuable of a technology to not take advantage of.

**Future application: screening for long noncoding RNA**

Only 1% of the mammalian genome carries protein coding potential, while 70–90% is transcribed at some point in development, producing an extensive and largely unexplored transcriptome of long noncoding
RNA (lncRNA) [69]. Many lncRNA are thought to have cis-regulatory roles such as recruitment of Polycomb repressive complexes or antisense regulation of neighboring genes [69]. Overexpression of lncRNA from conventional exogenous cDNA-based vectors will not be able to recapitulate potential cis-regulatory biological functions, intrinsically encoded by physical location of expression within chromatin.

Cas9-based epigenetic regulators, however, overcome this hurdle by altering expression directly from a target gene’s native locus. Thus, with the recognition that lncRNA are as important as protein coding genes in the human genome [70], it will be imperative to Cas9-based epigenetic regulators to perturb their functions in gain of function and loss of function screens, to uncover the full range of regulatory functions carried out by this poorly understood class of genes.

Future application: transcriptional and epigenetic gene therapy

There are several diseases that stand to benefit significantly from the application of Cas9 transcriptional regulators. Diseases such as Prader–Willi and Angelman Syndrome are caused by a pathogenic deletion in one parental allele, while the other fully functional allele is silenced due to genomic imprinting (the process by which certain alleles are silenced, depending on which parent it was inherited from). In such cases, it is possible that Cas9-based activators could activate expression from the silenced allele, providing expression of a fully intact copy of the missing gene.

Similarly, many monogenic diseases could be treated via modest overexpression of compensatory genes. For example, Duchenne muscular dystrophy (DMD) is an X-linked disease caused by mutations in the protein dystrophin. Pathogenic mutations disrupt the protein dystrophin, which is responsible for anchoring a cell’s actin cytoskeleton to the extracellular matrix in muscle. As dystrophin is located on the X chromosome, male patients only have a single copy of dystrophin; when that copy is mutated, no functional dystrophin is available, and muscle tissue loses its integrity causing muscle weakness and wasting. As a result, patients have difficulty with walking, respiration, and other daily activities. Currently, there are no commercialized therapies for DMD, but foreseen methods for treatment include the direct editing of genomic mutations in dystrophin to correct the faulty sequence [71,72].

An alternative approach, which has already been applied in clinical trials, is the up-regulation of a dystrophin paralog, utrophin, which can complement the function of mutated dystrophin [73]. In clinical trials, small molecule-mediated activation of expression is modest in its efficacy; direct activation of endogenous utrophin expression with CRISPR activators may solve this issue in a straightforward manner. Activators may be delivered episomally via AAV to muscle tissue, such as is the case with Glybera (the only approved gene therapy in the Western world, used to deliver a therapeutic gene to treat lipoprotein lipase deficiency) [74]. While such a method may not be a permanent cure, as the activator would not be stably integrated into the genome of target cells, it is safer than Cas9-based genome editing approaches that suffer from greater off-target effects, which are irreversible. Akin to Glybera, such a therapeutic may be delivered via local injection of gene therapy virus, and benefit from relatively long-term expression in muscle tissue without the risk of permanent integration [74].

Expanding upon this paradigm, many monogenetic disease genes have compensatory genetic paralogs that may be activated in a similar therapeutic approach. The highly programmable nature of Cas9 regulatory tools will allow for therapies developed for a single disease to be easily repurposed for other diseases that can be treated via activation or suppression of naturally encoded beneficial genes.

Concluding remarks

Clustered regularly interspaced short palindromic repeats and Cas9 based tools have filled the need for an efficient, programmable, user-friendly platform for genome editing as well as epigenetic regulation. As such, the field is moving at a rapid pace and quickly uncovering novel ways to adapt the system, as well as the rules for maximizing the efficacy of these tools. These tools have opened up the possibilities of rapid large-scale genome-wide screens for both gain of function and loss of function perturbations, mapping out functional networks with unprecedented depth and precision. While several aspects of CRISPR-based transcriptional and epigenetic regulators remain to be characterized, such as relative strengths of these tools, exact level of off-target activity, and ability to deliver these tools in vivo—we foresee many exciting applications such as probing of new classes of noncoding gene functions by regulating expression at endogenous loci, and applications to new paradigms of human gene therapy.

Acknowledgements

We thank Alejandro Chavez, Navpreet Ranu, Samouil Farhi, Justin Feng, Flo Yeo, and Amy Guo for their
assistance in editing the manuscript and critical discussions. G.M.C. is an inventor on issued patents (US9023649 and US9074199) filed by Harvard University related to CRISPR. G.M.C. is a founder and scientific advisor of Editas Medicine. G.M.C. has equity in Caribou/Intellia, Egenesis, and Editas (for full disclosure list, see: http://arep.med.harvard.edu/gmc/tech.html). S.V. acknowledges funding by the National Science Foundation Graduate Research Fellowship Program, the Department of Biological Engineering at the Massachusetts Institute of Technology, and the Department of Genetics at Harvard Medical School.

**Author contributions**

SV wrote the manuscript with support from all authors. MT collected and summarized data for comparison of tools. JC designed figure graphics and timelines in the free and open source software Inkscape. GC supervised the writing of the review.

**References**

1. Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD & Blancafort P (2012) Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* 7, 350–360.
2. Snowden AW, Gregory PD, Case CC & Pabo CO (2002) Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Curr Biol* 12, 2159–2166.
3. Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK & Bernstein BE (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* 31, 1133–1136.
4. Konermann S, Brigham MD, Trevino A, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM & Zhang F (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472–476.
5. Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD & Joung JK (2013) Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods* 10, 243–245.
6. Perez-Pinera P, Ousterout DG, Brunger JM, Farin AM, Glass KA, Gulik F, Crawford GE, Hartemink AJ & Gersbach CA (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods* 10, 239–242.
7. Bernstein DL, Le Lay JE, Ruanon EG & Kaestner KH (2015) TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J Clin Invest* 125, 1998–2006.
8. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE et al. (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* 31, 1137–1142.
9. Dhanasekaran M, Negi S & Sugiura Y (2006) Designer zinc finger proteins: tools for creating artificial DNA-binding functional proteins. *Acc Chem Res* 39, 45–52.
10. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA et al. (2008) Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31, 294–301.
11. Reyon D, Tsai SQ, Khayer C, Foden JA, Sander JD & Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30, 460–465.
12. Bolotin A, Quinquis B, Sorokin A & Ehrlich SD (2005) Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561.
13. Mojica FJM, Díez-Villaseñor C, García-Martínez J & Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60, 174–182.
14. Pourcel C, Salvignol G & Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653–663.
15. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA & Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.
16. Tang T-H, Bachellerie J-P, Rozhdestvensky T, Bortolin M-L, Huber H, Drungowski M, Elge T, Brosius J & Hüttenhofer A (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeglobus fulgidus*. *Proc Natl Acad Sci USA* 99, 7536–7541.
17. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV & van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964.
18. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13, 722–736.

S. Vora et al.
RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821.

20 Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

21 Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE & Church GM (2013) RNA-guided human genome engineering via Cas9. Science 339, 823–826.

22 Jinek M, East A, Cheng A, Lin S, Ma E & Doudna J (2013) RNA-programmed genome editing in human cells. Elife 2, e00471.

23 Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P & Moineau S (2008) Phase response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 190, 1390–1400.

24 Mojica FJM, Díez-Villaseñor C, García-Martínez J & Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740.

25 Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA & Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308.

26 Kosuri S & Church GM (2014) Large-scale de novo DNA synthesis: technologies and applications. Nat Methods 11, 499–507.

27 CustomArray, Inc. – maker of custom microarrays, oligo pools and instrumentation.

28 Sander JD & Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32, 347–355.

29 Weeks DP, Spalding MH & Yang B (2015) Use of designer nucleases for targeted gene and genome editing in plants. Plant Biotechnol J 13, 483–495.

30 Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L & Church GM (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31, 833–838.

31 Maeder ML, Linder SJ, Casco VM, Fu Y, Ho QH & Joung JK (2013) CRISPR RNA-guided activation of endogenous human genes. Nat Methods 10, 977–979.

32 Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW et al. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods 10, 973–976.

33 Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Weissman JS et al. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451.

34 Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivaila CS, Dadon DB & Jaenisch R (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res 23, 1163–1171.

35 Balboa D, Weltner J, Eurola S, Trokovic R, Wartiovaara K & Otonkoski T (2015) Conditionally stabilized dCas9 activator for controlling gene expression in human cell reprogramming and differentiation. Stem Cell Reports 5, 448–459.

36 Chakraborty S, Ji H, Kabadi AM, Gersbach CA, Christoforou N & Leong KW (2014) A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Reports 3, 940–947.

37 Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS & Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell 159, 635–646.

38 Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, Iyer EPR, Lin S, Kiani S, Guzman CD, Wiegand DJ et al. (2015) Highly efficient Cas9-mediated transcriptional programming. Nat Methods 12, 326–328.

39 Gilbert LA, Horibeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimarães C, Panning B, Ploegh HL, Bassik MC et al. (2014) Genome-Scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661.

40 Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H et al. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588.

41 Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS et al. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160, 339–350.

42 Hilton JB, D’Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE & Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33, 510–517.

43 Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP & Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183.

44 Thakore PI, D’Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, Reddy TE, Crawford GE & Gersbach CA (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods 12, 1143–1149.

45 Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M & Maehr R (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods 12, 401–403.
Kiani S, Chavez A, Tuttle M, Hall RN, Chari R, Ter-Ovanesyan D, Qian J, Pruitt BW, Beal J, Vora S et al. (2015) Cas9 gRNA engineering for genome editing, activation and repression. Nat Methods 12, 1051–1054.

Dahlman JE, Abudayyeh OO, Joung J, Gootenberg JS, Zhang F & Konermann S (2015) Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. Nat Biotechnol 33, 1159–1161.

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F & Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res doi: 10.1093/nar/gkt520.

Sternberg SH, LaFrance B, Kaplan M & Doudna JA (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 527, 110–113.

Platt RJ, Chen S, Zhou Y, Yim MJ, Sweich L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M et al. (2014) CRISPR-Cas9 Knockin mice for genome editing and cancer modeling. Cell 159, 440–455.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS et al. (2015) In vivo genome editing using staphylococcus aureus Cas9. Nature 520, 186–191.

Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S et al. (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol 32, 670–676.

Polstein LR, Perez-Pinera P, Kocak DD, Vockley CM, Bledsoe P, Song L, Safi A, Crawford GE, Reddy TE & Gersbach CA (2015) Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. Genome Res 25, 1158–1169.

The difficulty of a fair comparison. Nat Methods 12, 273–273.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O et al. (2013) DNA targeting specificity of RNA-guided Cas9 nuclease. Nat Biotechnol 31, 827–832.

Doench JG, Hartenian E, Graham DB, Tootova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ & Root DE (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol 32, 1262–1267.

Chari R, Mali P, Moosburner M & Church GM (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Methods 12, 823–826.

Wong N, Liu W & Wang X (2015) WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol 16, 218.

Zincarelli C, Solty S, Rengo G & Rabinowitz JE (2008) Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. Mol Ther 16, 1073–1080.

Gaudet D, de WJ, Tremblay K, Déry S, van DS, Freidig A, Brisson D & Métot J (2010) Review of the clinical development of alipogene tiparvovec gene therapy for lipoprotein lipase deficiency. Atheroscler Suppl 11, 55–60.

Mingozi F & High KA (2011) Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 12, 341–355.

Sweich L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M & Zhang F (2015) In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. Nat Biotechnol 33, 102–106.

Esvelt KM, Mali P, Braff JL, Moosburner M, Yang SJ & Church GM (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods 10, 1116–1121.

Kearns NA, Genga RMJ, Enamneh MS, Garber M, Wolfe SA & Maehr R (2014) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development 141, 219–223.

Gao X, Tsang JCH, Gaba F, Wu D, Lu L & Liu P (2014) Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. Nucleic Acids Res 42, e155–e155.

Limsirichai P, Gaj T & Schaffer DV (2016) CRISPR-mediated activation of latent HIV-1 expression. Mol Ther 24, 499–507.

Zhang Y, Yin C, Zhang T, Li F, Yang W, Kaminski R, Fagan PR, Putatunda R, Young W-B, Khalili K et al. (2015) CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. Sci Rep 5, 16277.

Lin S, Ewen-Campen B, Ni X, Housden BE & Perrimon N (2015) In vivo transcriptional activation using CRISPR-Cas9 in drosophila. Genetics 201, 433–442.

Lee JT (2012) Epigenetic regulation by long noncoding RNAs. Science 338, 1435–1439.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22, 1775–1789.

Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Rivera RMC, Madhavan S, Pan X, Ran FA, Yan WX et al. (2016) In vivo genome editing
improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403–407.

72 Tabebordbar M, Zhu K, Cheng JKW, Chew WL, Widrick JJ, Yan WX, Maesner C, Wu EY, Xiao R, Ran FA et al. (2016) In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407–411.

73 Wojtal D, Kemaladewi DU, Malam Z, Abdullah S, Wong TWY, Hyatt E, Baghestani Z, Pereira S, Stavropoulos J, Mouly V et al. (2016) Spell checking nature: versatility of CRISPR/Cas9 for developing treatments for inherited disorders. *Am J Hum Genet* **98**, 90–101.

74 Glybera, INN-alipogene tiparvovec. Annex I. Summary of Product Characteristics.