Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation

Antonia A. Dominguez1–3, Wendell A. Lim4–7 and Lei S. Qi1–5

Abstract | The bacterial CRISPR–Cas9 system has emerged as a multifunctional platform for sequence-specific regulation of gene expression. This Review describes the development of technologies based on nuclease-deactivated Cas9, termed dCas9, for RNA-guided genomic transcription regulation, both by repression through CRISPR interference (CRISPRi) and by activation through CRISPR activation (CRISPRa). We highlight different uses in diverse organisms, including bacterial and eukaryotic cells, and summarize current applications of harnessing CRISPR–dCas9 for multiplexed, inducible gene regulation, genome-wide screens and cell fate engineering. We also provide a perspective on future developments of the technology and its applications in biomedical research and clinical studies.

Complex and dynamic transcription regulation of multiple genes and their pathways drives many essential cellular activities, including genome replication and repair, cell division and differentiation, and disease progression and inheritance. Understanding the complex functions of a gene network requires the ability to precisely manipulate and perturb expression of the desired genes by repression or activation. However, until recently, we lacked such simple, robust technologies. RNA-mediated interference (RNAi), which uses small interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs), has been one major approach for sequence-specific gene suppression in eukaryotic organisms1. Although RNAi is a convenient tool for studying gene function, allowing transcript-specific degradation through Watson–Crick base-pairing through Watson–Crick base-pairing between mRNAs and siRNAs or shRNAs, its effects can be inefficient and nonspecific2. In addition to RNAi, customized DNA-binding proteins such as zinc-finger proteins or transcription activator-like effectors (TALEs) have been used as tools for sequence-specific DNA targeting and gene regulation3. These proteins robustly target DNA through programmable DNA-binding domains and can recruit effectors for transcription repression or activation in a modular way4–9. However, because each DNA-binding protein needs to be individually designed, their construction and delivery for the purpose of simultaneously regulating multiple loci is technically challenging10. Methods for gene overexpression include the use of cDNA overexpression vectors or vector libraries, but cloning large cDNA sequences into viral vectors and manipulating several gene isoforms simultaneously is difficult, and synthesizing large-scale libraries is costly. An ideal technology for genome regulation would therefore combine the convenience and scalability of RNAi with the robustness and modularity of DNA-binding proteins.

The discovery of the bacterial CRISPR–Cas system has inspired the development of a new approach for nucleotide base-pairing-mediated DNA targeting. The type II CRISPR system uses an endonuclease, Cas9, which is guided by a single guide RNA (sgRNA) that specifically hybridizes and induces a double-stranded break (DSB) at complementary genomic sequences11–14. Using an engineered nuclease-deficient Cas9, termed dCas9, enables the repurposing of the system for targeting genomic DNA without cleaving it15. As detailed below, recent work has suggested that dCas9 is a flexible, RNA-guided DNA recognition platform, which enables precise, scalable and robust RNA-guided transcription regulation.

In this Review, we first provide a very brief overview of the CRISPR–Cas9 technology for genome editing, before focusing on the development of CRISPR–dCas9 tools for transcription activation and repression in diverse organisms. We highlight the advantages and limitations of the current dCas9 technology, and also present a sampling of current applications of the technology in biological research and potential future clinical studies.

From editing to transcription control | CRISPR–Cas is an RNA-mediated adaptive immune system found in bacteria and archaea, in which it protects host cells from invasion by foreign DNA elements16.
**CRISPR–Cas** (Clustered regularly interspaced short palindromic repeats—CRISPR-associated proteins) CRISPR are bacterial DNA loci containing short repeat segments that match foreign DNA elements. Together with Cas proteins, they form an adaptive immune system in bacteria and archaea, which can acquire sequence segments from foreign DNA and use these sequences to recognize and destroy the foreign target DNA.

**Type II CRISPR system** CRISPR–Cas system that encodes cas9, cas1 and cas2 within the CRISPR–cos loci, in addition to a tracrRNA, which is partially complementary to repeats in the CRISPR array.

**Single guide RNA (sgRNA)** A synthetic RNA chimeras containing a hairpin that links the transactivating CRISPR RNA (tracrRNA) to the crRNA and functions similarly to the native crRNA–tracrRNA duplex, directing Cas9 to a specific genomic locus.

**Protospacer-adjacent motif (PAM)** A short sequence in the target DNA (not in the guide RNA) that is necessary for the successful targeting of Cas9. The PAM sequence varies between bacterial species. In *Streptococcus pyogenes*, it is NGG (which is more effective) or NAG (less effective).

**Transactivating CRISPR RNA (tracrRNA)** A small RNA encoded upstream of the CRISPR locus in type II CRISPR systems, with a 24-nucleotide sequence complementary to repeats of the CRISPR RNA (crRNA) precursor transcripts. Essential for the processing of pre-crRNA to mature crRNA.

**CRISPR RNA (crRNA)** Small RNAs transcribed from the protospacer within the CRISPR array. Together with the transactivating crRNA (tracrRNA), crRNA guides Cas9 to a specific genomic locus.

CRISPR–Cas is currently divided into two major classes and five types, of which type II is the most widely used for genome-engineering applications. Discovery of key components of the type II CRISPR system and elucidation of its mechanism were integral to its use as a genome-engineering tool. These include the demonstration that *Streptococcus thermophilus* could specifically cleave double-stranded DNA, mediated by Cas9 (REFS 11, 12); the discovery of a short DNA sequence adjacent to the RNA-binding site, later termed the protospacer-adjacent motif (PAM), as the CRISPR–Cas mechanism for discriminating self from non-self; the discovery of a small transactivating CRISPR RNA (tracrRNA), which directs the post-transcriptional processing and maturation of the CRISPR RNA (crRNA) through sequence complementarity; and, lastly, the demonstration that the CRISPR–Cas system from *S. thermophilus* could function in *Escherichia coli* and provide resistance against foreign plasmids. On the basis of these findings about CRISPR–Cas9 biology, it was demonstrated that the *Streptococcus pyogenes* Cas9 protein could bind to a tracrRNA–crRNA complex or to a designed, chimeric sgRNA to generate a double-strand break (DSB) at a specific site of the target DNA in *vitro*. Another report similarly showed that *S. thermophilus* Cas9 could interact with the tracrRNA–crRNA complex to cut DNA. Demonstrations of the use of Cas9 and sgRNAs for genome editing in *vitro* rapidly followed this seminal observation. Further information on the genome-editing applications of CRISPR–Cas9 can be found in other reviews.

In addition to using the nuclease Cas9 for editing genomic sequences, the CRISPR–Cas9 technology can be used as a sequence-specific, non-mutagenic gene regulation tool. This repurposing was first demonstrated by introducing mutations into the *S. pyogenes* Cas9 in its two nuclease domains, HNH and RuvC. The resulting nuclease-deficient dCas9 is unable to cleave DNA but retains the ability to specifically bind to DNA when guided by a sgRNA. As discussed below, dCas9 allows for direct manipulation of the transcription process without genetically altering the DNA sequence. Furthermore, it allows the recruitment of diverse effector proteins for gene regulation at the transcription level. Other uses of the dCas9 protein include chromosome imaging in live cells and dissection of long-range chromatin interactions.

**Transcription repression by CRISPRi** Bacteria lack the machinery for RNAi, and simple platforms for targeted gene regulation in bacteria have been limited. The utility of dCas9 for sequence-specific gene repression was first demonstrated in *E. coli* as a technology called CRISPR interference (CRISPRi). By pairing dCas9 with a sequence-specific sgRNA, the dCas9–sgRNA complex can interfere with transcription elongation by blocking RNA polymerase (Pol). It can also impede transcription initiation by disrupting transcription factor binding. In bacteria, the CRISPRi method using dCas9 is highly efficient in suppressing genes; is specific, with minimal off-target effects; and is multiplexable, such that several genes can be simultaneously controlled using multiple sgRNAs. Unlike the permanent genetic modifications induced by the nuclease Cas9, gene repression using CRISPRi is reversible. A disadvantage is that dCas9 may repress downstream genes within an operon (polar effects) instead of an individual gene. The CRISPRi platform thus provides a robust RNA-guided approach for gene regulation.
The introduction of CRISPRi into mammalian cells using dCas9 alone achieved only modest repression of enhanced GFP (egfp) in the human HEK293T reporter cell line\(^\text{15}\). When targeting endogenous genes such as the transferrin receptor CD71, C-X-C chemokine receptor type 4 (CXC4) and tumour protein 53 (TP53), up to 80% repression was observed\(^\text{12,14}\). To achieve enhanced repression, the Krüppel-associated box (KRAB) or four concatenated mSin3 interaction domains (SID4X) was fused to the carboxyl terminus of dCas9. Together with a target-specific sgRNA, the dCas9–KRAB or dCas9–SID4X fusion proteins can efficiently repress endogenous genes (CXC4, CD71, Krüppel-like factor 4 (KLF4) or SRY-box 2 (SOX2)) in mammalian cells\(^\text{31–40}\) (FIG. 2a). This repression was further enhanced by fusing KRAB to the amino terminus of dCas9, leading to strong repression of endogenous genes\(^\text{41}\). The level of dCas9– KRAB–dCas9-mediated knockdown of endogenous genes was highly dependent on the sgRNA targeting site, suggesting that the chromatin structure or the presence of regulatory elements may limit the level of repression. In yeast, a different mammalian transcription repressor domain, Max-interacting protein 1 (Mxi1), was used for effective repression\(^\text{42}\). CRISPRi has been used in genome-wide screens and for the manipulation of cell fate, which are discussed below.

Transcription activation by CRISPRa. CRISPR-mediated gene activation, termed CRISPRa, uses dCas9 fusion proteins to recruit transcription activators. A fusion of dCas9 with the ω-subunit of the E. coli Pol allowed assembly of the holoenzyme at a target promoter for gene activation in E. coli\(^\text{43}\). There are currently limited reports on CRISPRa in bacteria, and more work is needed to achieve robust and consistent gene activation in bacteria.

The fusion of VP64 or of the p65 activation domain (p65AD) to dCas9 in mammalian cells could activate both reporter genes and endogenous genes, with a single sgRNA\(^\text{38,42–44}\) (FIG. 2b). However, the use of multiple sgRNAs was necessary to achieve significant activation of the endogenous genes tested (interleukin 1 receptor antagonist (IL1RN), achaete-scute family bHLH transcription factor 1 (ASCL1), Nanog homeobox (NANOG), myogenic differentiation 1 (MYOD1), vascular endothelial growth factor A (VEGFA) and neurotrophin 3 (NTF3))\(^\text{44}\). Protein engineering approaches were adopted to optimize the efficiency of activation. For example, it was determined that stronger activation could be achieved with VP64 fused simultaneously at both amino and carboxyl termini\(^\text{44}\). The addition of multiple copies of VP16 (for example, dCas9–VP160) was reported; however, the efficient activation of endogenous IL1RN, octamer-binding 4 (OCT4; also known as POU5F1) and SOX2 still required multiple sgRNAs\(^\text{45}\).

The complexity of genome-wide activation screens or cell fate reprogramming experiments necessitate the use of one efficient sgRNA per gene. The enhancement of gene activation observed with multiple sgRNAs suggested that recruitment of many activators could increase activation efficiency. One study used dCas9
fused with a carboxy-terminal SunTag array, which consisted of 10 copies of a small peptide epitope35. A cognate single-chain variable fragment (scFV) fused to a superfolder GFP (sfGFP; for improving protein folding) and to VP64 (scFV–sfGFP–VP64) recognized these peptides and recruited multiple copies of VP64 to a single dCas9. Using dCas9–SunTag, significant activation of CXCR4 was achieved with a single sgRNA, leading to the modulation of cell migration35 (FIG. 2c). An additional study screened different activator domains and members of the Mediator complex and Pol II complex for highly efficient activation of endogenous genes. The screen led to the development of a tripartite activator domain that consisted of VP64, p65AD and the Epstein–Barr virus R transactivator Rta47 (VPR) (FIG. 2c). The dCas9–VPR fusion showed improved activation of endogenous coding and non-coding genes using multiple sgRNAs when compared with dCas9–VP64. The system was also tested in Saccharomyces cerevisiae, Drosophila melanogaster and Mus musculus cells for activating endogenous loci47.
The ability to target epigenetic modifications to a gene in a combinatorial fashion may allow the temporal and spatial regulation of genes that are natively regulated by a complex set of interacting transcription factors. In addition to dCas9 engineering, sgRNA engineering was also shown to enhance the efficiency of gene activation. The recruitment of VP64 using protein-interacting RNA aptamers incorporated into the sgRNA has achieved activation of the gene encoding endogenous zinc-finger protein 42, using multiple sgRNAs. An improvement, termed the synergistic activation mediator (SAM) system, was achieved by adding MS2 aptamers to the sgRNA; MS2 recruits its cognate MS2 coat protein (MCP) fused to p65AD and heat shock factor 1 (HSF1). Additionally, VP64 is fused to dCas9. Epigenetic regulation can be carried out by fusion of epigenetic regulators to dCas9. Fusion of the histone demethylase LSD1 to Neisseria meningitidis Cas9 removes the histone 3 Lys4 dimethylation (H3K4me2) mark from target distal enhancers, leading to transcription repression. The fusion of the catalytic core of the histone acetyltransferase p300 (p300<sup>ΔC</sup>) to dCas9 can acetylate H3K27 (H3K27ac) at targeted proximal and distal enhancers, which leads to transcription activation.

**dCas9-mediated epigenetic modifications.** The ability to manipulate epigenetic modifications, such as histone acetylation and methylation and DNA methylation, would allow for the interrogation of epigenetic regulation of cellular function. The histone demethylase LSD1 (Lys-specific histone demethylase 1) fused to Neisseria meningitidis dCas9 was recently used for gene repression<sup>49</sup>. Using dCas9–LSD1 and a sgRNA in mouse embryonic stem cells (ES cells) to target the distal enhancer region of the endogenous transcription factor gene Oct4, the authors demonstrated the repression of Oct4 and loss of pluripotency. However, downregulation of Oct4 expression was not seen when the complex targeted the proximal enhancer region, which is known to regulate Oct4 expression in epiblast cells<sup>52</sup>. This indicates that this epigenetic regulatory system can allow delineation between cell type-specific enhancers (Fig. 2e).

Additionally, rather than catalysing a specific histone modification, it was recently demonstrated that targeting dCas9–KRAB with a sgRNA to the HS2 enhancer in the globin locus control region led to H3K9 trimethylation at the enhancer, thus silencing the expression of multiple globin genes<sup>51</sup>. The use of these tools to silence transcription by targeting regulatory regions, instead of the target gene itself, further expands the capacity of dCas9 as a versatile transcription manipulation tool.

In addition to utilizing activation domains to achieve endogenous gene activation, the catalytic core of the human acetyltransferase p300 was recently fused to dCas9 (Cas9–p300<sup>ΔC</sup>) for targeted epigenetic regulation. Target genes were activated by catalysing the acetylation of histone H3 Lys27 (H3K27ac) at both promoters and enhancers<sup>53</sup>. Although potential off-target binding may lead to spurious activation, owing to the possibility of activating distant enhancers, dCas9–p300<sup>ΔC</sup> was found to be specific and robust, only activating the targeted gene (Fig. 2f). Taken together, these studies demonstrate that dCas9 fused to epigenetic modifiers can modulate chromatin states and gene expression, thereby providing powerful tools for probing the interactions between the epigenome, regulatory elements and gene expression. The ability to target epigenetic modifications to a gene in a combinatorial fashion may allow the temporal and spatial regulation of genes that are natively regulated by a complex set of interacting transcription factors<sup>44</sup>.

**Simultaneous activation and repression.** CRISPR–dCas9 can target several genes simultaneously by using multiple sgRNAs. Recently, a method for simultaneous repression and activation of genes was established using scaffold RNAs (scRNAs)<sup>54</sup>. The scRNAs are designed by extending the sgRNA sequence with orthogonally acting protein-binding RNA aptamers (MS2, PP7 or com)<sup>55</sup>. Each scRNA can encode information both for DNA target recognition and for recruiting a specific repressor or activator protein. By changing the DNA targeting sequence or the RNA aptamers in a modular fashion, multiple dCas9–scRNAs can simultaneously activate or repress multiple genes in the same cell (Fig. 3). This functionality could facilitate the study of regulatory networks and genetic interactions. For example, a scRNA-based strategy was developed to modulate a branched metabolic pathway in yeast cells<sup>56</sup>, wherein different combinations of scRNAs were used to activate and repress alternative sets of enzymes for the production of distinct metabolites. In mammalian cells, two scRNAs were used to simultaneously activate CXCR4 with two MS2 scRNAs recruiting VP64, and repress β-1,4-N-acetylgalactosaminyl transferase 1 (B4GALNT1) with a com scRNA recruiting KRAB (Fig. 3).

The diversity of CRISPR–Cas systems highlights the potential of using orthogonal dCas9 proteins for parallel gene regulation<sup>54</sup>. Two limiting factors for this approach are the design of functional cognate sgRNAs and the characterization of the full PAM landscape. In addition...
Figure 3 | Simultaneous transcription activation and repression. Multiplexed transcription activation and repression is carried out using single guide RNAs (sgRNAs) modified with RNA aptamers, termed scaffold RNAs (scRNAs). The com aptamer recruits the aptamer-binding protein, Com, which was fused to a Krüppel-associated box (KRAB) domain for transcription repression of the aptamer-binding protein, Com, which was fused to a Krüppel-associated box (KRAB) modified with RNA aptamers, termed scaffold RNAs (scRNAs). The com aptamer recruits transcription activation and repression is carried out using single guide RNAs (sgRNAs) into a window from -50 to +300 bp relative to the TSS of a specific gene. The specificity of Cas9 and dCas9

The specificity of the nuclease Cas9 in mammalian cells remains a major concern for the use of the technology, in particular for clinical purposes. Compared to the bacterial cells in which CRISPR–Cas9 has evolved, the several-hundred-fold larger mammalian genomes might present many more off-target binding sites to the system. Thus, the off-target effects of CRISPR–Cas9 in genome-wide binding, editing and regulation have been examined extensively. It is important to distinguish the difference between the three cases (binding, editing and regulation), as off-target binding may not necessarily have editing or regulatory effects. To examine binding specificity, two studies mapped the genome-wide binding sites of dCas9 with multiple different sgRNAs in mouse ES cells and HEK293T cells. Chromatin immunoprecipitation followed by deep DNA sequencing (ChIP–seq) analysis revealed that dCas9 had bound to many off-target genomic sites. However, through targeted sequencing of the dCas9 binding sites, these studies demonstrated that cleavage by Cas9 at off-target sites was substantially lower than at on-target sites. This suggested that although the high level of off-target binding of Cas9 is a concern, only a small subset of off-target binding sites were cleaved efficiently.

In addition to binding, various approaches were used to characterize editing specificity. These studies suggest that off-target effects are a concern for gene editing in mammalian cells, but that they may be highly dependent on the target gene, the sequence of the designed sgRNA, the cell type, and the context of the genomic sequence and its epigenetic state.

Gene repression was found to be quite specific when the transcriptome of HEK293T cells expressing dCas9–KRAB with a targeting or a non-targeting sgRNA was assayed by RNA sequencing (RNA-seq). Similar results were reported when using dCas9–VP64 for targeted gene activation. In another study, it was observed that, on average, more than two sequence mismatches between the sgRNA and the target gene abolished CRISPRi regulatory activity in a set of genes, and therefore that dCas9 used for gene regulation can tolerate fewer mismatches compared with Cas9 used for gene editing. To test the concern that the fused effector domains might contribute additional off-target binding of dCas9, ChIP–seq was conducted using dCas9–KRAB or dCas9–VP64 with multiple sgRNAs. The results showed that genome-wide binding profiles were similar to those of dCas9 without an effector domain and were also highly specific.

It is thought that the specificity of dCas9 for gene regulation comes from the fact that effective transcription regulation requires dCas9–sgRNA to bind within a small ‘window’ of sequence around the transcription start site (TSS) and to interact with local transcription factors or Pol complexes. A set of rules was determined to optimize the efficacy of sgRNAs in modulating gene expression by testing a library of sgRNAs targeting the region surrounding the TSSs of 49 genes that had previously been shown to make cells susceptible to ricin. For CRISPRi, strong repression was achieved when KRAB–dCas9 was targeted to a window from -50 to +300 bp relative to the TSS of a specific gene, with maximum repression detected at +50 to +100 downstream of the TSS. In addition, sgRNAs with protospacer lengths of 18–21 bp were more active than those with longer protospacers, whereas a sequence of identical bases (such as TTTT or CCCC) in the sgRNAs had a negative effect on repression. Neither the choice of targeting strand nor the GC content of the sgRNAs correlated with repression levels. For CRISPRa using dCas9–SunTag, optimal sgRNA-mediated gene activation was found when targeting a window between -50 and -400 bp upstream of the TSS; for activation using the SAM system, the optimal window was determined to be between -200 and +1 bp relative to the TSS. Thus, most off-target binding events of dCas9, which occur outside these sequences, may not lead to changes in transcription. Furthermore, many off-target binding events may be transient and therefore insufficient for modulating
transcription of nearby genes. Although the current data in mammalian cells demonstrate that the CRISPRi and CRISPRa systems are specific, further studies are needed to fully understand the causes of off-target binding and to develop more strategies to minimize it.

CRISPR–dCas9 applications

The CRISPR–dCas9 system is a broadly applicable tool for genome-scale screening, manipulation of dynamic gene programmes and modulation of cell fates. Here, we describe these applications and compare them to alternative approaches.

Genome-wide screens. The ability to regulate essentially any genomic locus enables the study of gene function on a global scale. RNAi has been used for genome-wide screens; however, concerns about its efficiency and specificity still remain\(^1\). Overexpression screening methods have relied on the construction and delivery of cDNA vector libraries; however, difficulties exist in manipulating multiple gene isoforms simultaneously, in addition to the high cost and difficulty of cloning such cDNA libraries. The ability to easily design and clone sgRNAs makes the Cas9 system a powerful approach for genome-wide screens using oligonucleotide synthesis. Several studies have used Cas9 to conduct genetic knockouts for genome-scale loss-of-function screens\(^2\)–\(^5\).

Distinct from Cas9-mediated screens, the dCas9 systems allow for both genome-wide loss-of-function (using CRISPRi) and gain-of-function (using CRISPRa) screens. The CRISPRi and CRISPRa screens are based on the pooled approach, in which sgRNAs are synthesized as a mixture of oligonucleotides and then cloned in mixture to generate an sgRNA vector library (FIG. 4a). This library is packaged into viral particles that are used to transduce mammalian cells at a low multiplicity of infection, achieving genomically integrated rates of one sgRNA per cell. The different sgRNAs are barcoded so that their identity can be assayed by deep sequencing to infer which gene is targeted, and thus activated or repressed, in any particular cell. The relative abundance of each sgRNA at the end of the screen is indicative of the effect of silencing or activating the targeted gene under the specific experimental conditions (FIG. 4a).

A recent study used a genome-wide sgRNA library targeting each gene with ten sgRNAs per gene for CRISPRi screening. Myelogenous leukaemia K562 cells expressing KRAB–dCas9 and the sgRNA library were cultured with or without a chimeric toxin composed of the diphertheria toxin catalytic subunit linked to cholera toxin (CTX–DTA)\(^6\). The screen revealed both known and unanticipated genes that control sensitivity to CTX–DTA\(^6\). Additionally, using a large library of nontargeting sgRNAs, the researchers found that 99.5% of control sgRNAs had no activity, thus demonstrating the high specificity of the CRISPRi system. Robust repression (80–99% knockdown of genes) was demonstrated by validating the top hits individually. The strong repression and low off-target activity are clear advantages of CRISPRi; however, there are still important uses for RNAi-based screens. For example, CRISPRi modulates transcription at the TSSs of endogenous genes; therefore, it is difficult to target specific splice isoforms. By contrast, RNAi can be targeted to specific mature transcripts\(^4\). Therefore, the use of CRISPRi and RNAi in conjunction may hold the potential for more complete analysis of gene function.

Two reports have demonstrated the use of CRISPRa for genome-wide screens. As a complementary approach to the CRISPRi screen, one CRISPRa screen utilized the dCas9–SunTag system to probe genes that modulate sensitivity to CTX–DTA\(^4\). Interestingly, this gain-of-function screen provided both new and complementary information to the results of the CTX–DTA CRISPRi screen. In another study, the SAM system was used to activate all human transcript isoforms in a malignant melanoma cell line and screen for genes that confer resistance to an inhibitor of the proto-oncogene Ser/Thr kinase B-RAF (BRAF)\(^9\). They discovered novel resistance-conferring candidates, in addition to validating known resistance genes. The advantages of CRISPRa screens over the cDNA overexpression approach include the ability to assay the consequences of activating an endogenous gene locus, and the ability to drive the expression of multiple splicing isoforms with one targeting sgRNA\(^8\). The possibility of manipulating multiple genes in single cells may enable large-scale screens that will help to elucidate genetic interactions and uncover networks of proteins that are important for cell fate and function.

Gene regulation by inducible CRISPR–dCas9 tools.

CRISPR–dCas9 can be combined with other tools to control gene expression in a spatial and temporal manner, which is useful for understanding dynamic gene networks. Several transcription control strategies based on optogenetics have been developed that utilize light-inducible peptide heterodimerization, in which one peptide is fused to a DNA-binding protein and another to a transcription activator. Two groups recently created such light-activated dCas9-effectors using the cryptochrome-based blue light-sensing system CRY–CIB heterodimerizing domains to recruit VP64 or p65AD to dCas9 (REFS 75, 76) (FIG. 4b). By illuminating cells with blue light, both studies demonstrated the activation of endogenous genes using a mixture of four sgRNAs; the highest activation levels were comparable to those obtained with the dCas9–VP64 system in HEK293 cells. Although using single sgRNAs with these systems resulted in poor gene activation, these studies demonstrated that targeted gene regulation could be spatially and temporally controlled in a reversible manner using light. In addition to optical induction, a chemically-inducible system for activating endogenous loci has also been developed on the basis of rapamycin-dependent dimerization of a split dCas9–VP64 (REF. 77) (FIG. 4c). In the future, discovery of other optogenetically inducible (for example, the recently reported nMag system\(^79\)) or chemically inducible dimerization systems may expand and optimize inducible dCas9-based regulation tools, which would then ideally be introduced into whole organisms to drive activation or repression of precise spatial and temporal gene expression programmes in vivo.

**Optogenetics**

A technique that utilizes optics for achieving spatiotemporal gene regulation of cells in living tissues.

**CRY–CIB heterodimerizing domains**

A light-inducible protein interaction between the blue light-sensitive cryptochrome 2 protein (CRY2) and its interacting partner CIB1 from *Arabidopsis thaliana*. 
Figure 4 | Applications of the CRISPR–dCas9 technology. a | Overview of pooled screening using CRISPR interference (CRISPRi) or CRISPR activation (CRISPRa). Growth-based screens identify targeting single guide RNAs (sgRNAs) that confer growth advantage or disadvantage on the basis of sgRNA enrichment or depletion in the final population, assessed using deep sequencing. b | Optogenetics-based transcription control using light-dependent peptide heterodimerization. The amino-terminal fragment of CIB1 (CIBN) is fused to both the amino and carboxyl termini of dCas9. The blue light-sensitive cryptochrome 2 protein CRY undergoes a conformational change in the presence of blue light that enables its heterodimerization with CIBN, thus recruiting activation domains (green dashed oval) such as the herpes simplex virus VP16 activation domain (VP64) or the nuclear factor-κB (NF-κB) transactivating subunit activation domain (p65AD). c | In this chemically inducible system for transcription activation, rapamycin induces the dimerization of FK506-binding protein (FKBP) and the FKB–rapamycin-binding domain (FRB). Thus, a dCas9 protein that is split into dCas9 (N terminus)–FRB and FKBP–dCas9 (C terminus)–VP64 can be reassembled by the introduction of rapamycin. d | The targeted activation of octamer-binding 4 (OCT4) with dCas9–VP192 (12 repeats of VP16) can replace transgenic expression of OCT4 to achieve reprogramming from a differentiated cell to a pluripotent stem cell, but this also requires the transgenic expression of SRY-box 2 (SOX2), Krüppel-like factor 4 (KLF4), LIN-28 homologue A (LIN28) and MYC, and the knockdown (KD) of p53. e | Fusion of two VP64 domains to dCas9 induced the expression of myogenic differentiation 1 (Myod1) in mouse embryonic fibroblasts, which caused direct cell reprogramming into skeletal myocytes. f | Activation of neurogenin 2 (NGN2) and neuronal differentiation 1 (NEUROD1), using dCas9–VPR (where VPR is a complex of VP64, p65AD and Epstein–Barr virus R transactivator Rta) and a mixed pool of sgRNAs, directed the differentiation of pluripotent stem cells into neuron-like cells.
**Cell fate engineering.** By specifically controlling gene expression, CRISPRi and CRISPRa can be used to modulate cell identity, reprogramming and differentiation. To reprogramme HEK293T cells into induced pluripotent stem cells (iPSCs), the human OCT4 promoter was activated by targeting of multiple sgRNAs with dCas9–VP64; only modest gene activation was achieved\(^8\), which was considerably enhanced by expression of the epigenetic modifier p300, albeit to levels that were still not sufficient to drive reprogramming. Although somatic cell reprogramming into human iPSCs solely using CRISPRa has not been achieved, dCas9–VP192 (12 copies of VP16) in combination with multiple sgRNAs targeting the OCT4 promoter was able to replace transgenic OCT4 expression, but reprogramming still required the overexpression of the additional reprogramming factors\(^8\) (FIG. 4a). In addition to somatic cell reprogramming into iPSCs, direct lineage reprogramming has been attempted using CRISPRa tools. A fusion of two VP64 domains flanking dCas9 was used to induce the transcription of Myod1 in mouse embryonic fibroblasts, which caused them to differentiate into skeletal myocytes\(^9\) (FIG. 4e). Activation of human MYOD1 was also achieved; however, levels were much lower than for mouse Myod1 and were not sufficient to reprogramme human fibroblasts to skeletal myocytes\(^9\). These results are promising for the use of CRISPRa in reprogramming; however, it is clear that current levels of activation are insufficient to drive reprogramming or direct lineage reprogramming of most human cell types.

Achieving robust and homogenous differentiation of pluripotent cells will be essential for disease modelling, and CRISPRa or CRISPRi have potential to direct such differentiation. The activation of a key marker of endoderm, SOX17, was achieved using dCas9–VP64 with multiple sgRNAs in human ES cells\(^9\). The same group then tested the ability of dCas9–KRAB to repress OCT4 in human ES cells. They achieved significant repression of OCT4, as well as downregulation of NANOG, influencing the pluripotency expression network\(^9\). Additionally, enhanced activation systems such as dCas9–VPR have been utilized for the differentiation of human iPSCs into neuronal cells by activating neurogenin 2 (NGN2) and neuronal differentiation 1 (NEUROD1) with a mixed pool of 30 targeting sgRNAs\(^7\) (FIG. 4f). These studies provide promising evidence of the ability to use CRISPRi or CRISPR for direct reprogramming and differentiation. This would provide a new, CRISPR-based approach for cell fate modulation, improving our ability to use pluripotent stem cells for disease studies and for future therapies.

**Conclusions and future perspectives**

The CRISPR–Cas9 nuclease system offers a powerful approach for precisely modifying genomic sequences, allowing the study of gene function at nucleotide resolution. The ability to correct genetic mutations in a permanent manner will be an important aspect of this tool for future therapeutics. Another advantage of the CRISPR–Cas9 nuclease is that it enables complete genetic loss of function. However, loss of function often results in bimodality, wherein a cell population exhibits loss of function, while other cells acquire in-frame mutations and may retain gene function. This can be partially alleviated by the use of homology-directed repair with CRISPR–Cas9 but, to date, this process remains inefficient. We believe that CRISPRi and CRISPRa are useful tools to use in concordance with gene-editing strategies. The nuclease-deactivated dCas9 offers the ability to transiently or stably control gene expression without altering the genomic sequence. Partial loss-of-function studies are important for our understanding of gene function, in particular when studying essential genes. In addition, this technology offers a relatively simple method for manipulating the expression of multiple genes and thus is also important for the study of polygenic diseases.

Furthermore, CRISPRi and CRISPRa may allow quantitative tuning of gene expression and thus understanding of how gene dosage drives processes such as cell proliferation and differentiation or disease progression. Studies to enhance the efficiency of repression or activation, the development of inducible tools, and the creation of improved orthogonal systems for parallel activation and repression in the same cell will further broaden the use of CRISPRi and CRISPRa.

The strength of the Cas9 nuclease system in studying human disease has been demonstrated by the correction of genetic mutations in animal models\(^10\–13\). Genomewide association studies have identified many disease- and trait-associated genetic variants, with up to 93% of these found outside the protein-coding sequence. This implies that the aberrant regulation of gene expression and non-coding RNAs is important in the aetiology of diseases\(^14\). Thus, methods for manipulating gene expression could be vital for disease research. The use of CRISPRi for in vivo gene regulation is likely to offer an alternative to RNAi for studying gene function, and for modelling and therapeutics. Perhaps even more significantly, in vivo activation studies are likely to benefit from CRISPRa, as activation of multiple genes can be achieved simply by expressing several small sgRNAs.

The CRISPRi and CRISPRa technologies will benefit from the discovery and use of other Cas9 orthologues, in addition to the creation of dCas9-knock-in animal models. The most commonly used S. pyogenes Cas9 protein is encoded by a 4.2 kb gene, which is just within the packaging limit of adeno-associated virus (AAV) vectors. Recently, a smaller Cas9 orthologue, from S. aureus, was shown to have similar editing capabilities to the S. pyogenes Cas9, but its gene is 25% shorter. The smaller size facilitates its packaging with a sgRNA cassette into a single AAV vector for in vivo delivery\(^15\). Other Cas9-like proteins, such as Cpf1 (CRISPR from Prevotella and Francisella 1), have been shown to exhibit different mechanisms for DNA cleavage\(^16\). It would be interesting to look at the nuclease-deactivated versions of these proteins and explore their potential for sequence-specific gene regulation. For example, they may show different binding affinities and/or interact with local transcription factors differently. In summary, there is much to be explored before we develop fully comprehensive CRISPR–dCas9 or dCas9-like toolkits for transcription regulation and related biomedical research and clinical applications.
REVIEWS


34. Chadwick, J. S. & Williams, D. G. CRISPR-Cas9 systems that target Dcr-2 and Crr9 genes have substantial off-target activity. Nucleic Acids Res. 41, 5958–5959 (2013).


Acknowledgements
The authors thank the members of the Qi lab for advice and helpful discussions. L.S.Q. acknowledges support from the U.S. National Institutes of Health (NIH) Office of the Director (OD) and National Institute of Dental & Craniofacial Research (NIDCR). A.A.D. acknowledges support through the Milton Safenowitz Post Doctoral Fellowship for ALS Research. This work was supported by NIH RO1 DA036858 to W.A.L. and L.S.Q., the Howard Hughes Medical Institute (grant to W.A.L) and NIH DP5 OD017887 (to A.A.D. and L.S.Q.).

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION

ALL LINKS ARE ACTIVE IN THE ONLINE PDF