dCas9: A Versatile Tool For Epigenome Editing

Daan J.W. Brocken¹, Mariliis Tark-Dame²* and Remus T. Dame¹,3*

¹Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, Leiden, The Netherlands.  
²Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands.  
³Centre for Microbial Cell Biology, Leiden University, Leiden, The Netherlands.  
*Correspondence: m.tark@uva.nl and rtdame@chem.leidenuniv.nl  
https://doi.org/10.21775/9781910190630.02

Abstract
The epigenome is a heritable layer of information not encoded in the DNA sequence of the genome, but in chemical modifications of DNA or histones. These chemical modifications, together with transcription factors, operate as spatiotemporal regulators of genome activity. Dissecting epigenome function requires controlled site-specific alteration of epigenetic information. Such control can be obtained using designed DNA-binding platforms associated with effector domains to function as targeted transcription factors or epigenetic modifiers. Here, we review the use of dCas9 as a novel and versatile tool for fundamental studies on epigenetic landscapes, chromatin structure and transcription regulation, and the potential of this approach in basic research in these fields.

Introduction
The epigenome is a layer of information that, together with transcription factors, defines the cell-type-specific gene expression pattern of a genome. By definition, epigenetic information is mitotically and/or meiotically heritable, but not directly encoded in the DNA sequence (Bird, 2002; Berger et al., 2009). Epigenetic information consists of covalent chemical modifications [post-translational modification of histone proteins that include, but are not limited to, methyl, acetyl, phosphoryl and ubiquitin groups (Turner, 2012) and methylation of cytosine bases (Vardimon et al., 1982; Schübeler, 2015)] that alter the structure and physicochemical properties of DNA or DNA-bound histones (Bird, 2002; Kouzarides, 2007). Epigenetic modification states are dynamic by nature and depend on enzymes transferring or removing these modifications (Holliday, 1987; Cubas et al., 1999; Chong and Whitelaw, 2004; Youngson and Whitelaw, 2008). The epigenetic state of a genomic region is determined by combinations of modifications, but how exactly the resulting code is determined remains poorly understood (Gardner et al., 2011). These states correlate with gene expression and chromatin structure, and link the modification pattern of DNA and histones of genomic regions to states of development and differentiation (Holliday, 2006; Henikoff and Shilatifard, 2011; Zhou et al., 2011; Turner, 2012; Smith and Meissner, 2013).

A variety of diseases has been linked to mutations in epigenetic maintenance enzymes or to misregulation of genes following aberrations in the epigenetic code (Kelly et al., 2010; Baylin and Jones, 2011; Plass et al., 2013). Epigenome editing could develop into a tool used to revert aberrations in this code that lead to disease. Due to the lack of knowledge of rare modifications and the combinatorics of DNA and histone modifications, it is unclear by what exact mechanisms epigenetic signals lead to downstream effects on gene regulation and chromatin conformation and whether as yet unknown
functional elements exist in genomes. To be able to provide answers to these questions, especially those on causality and the order of concerted processes, epigenetic effectors have been employed to study the implications of modification of the epigenome. Sequence-specific targeting of epigenetic modifiers has been instrumental in understanding the roles of epigenetic modifications on gene regulation. Here, we review current efforts aimed at manipulating the epigenome, focusing on the nuclease-deactivated Cas9 (dCas9) as versatile tool for sequence-specific recruitment of effector proteins.

Epigenome editing
Numerous studies have established correlations between epigenetic state and genome activity (Bernstein et al., 2010; Rivera and Ren, 2013; Roadmaps Epigenomics Consortium et al., 2015). Investigating the epigenome and establishing causal relationships rather than correlations will benefit from site-specific targeting of enzymes involved in establishing, disrupting and maintaining epigenetic states. Ideally, approaches to achieve this aim are readily adaptable to accommodate different target specificity and functionality.

The earliest efforts to edit epigenetic states and determine the effects of alterations therein involved gene knock-out or knock-down of specific epigenetic modifiers (Stancheva and Meehan, 2000; Webster et al., 2005), or the use of compounds inhibiting the activity of these modifiers or their targets (Lyko and Brown, 2005). Such global strategies, due to their pleiotropic impact, do not permit direct specific changes to be distinguished from secondary effects of the perturbation. Epigenetic modifiers – or their isolated functional domains – have been fused to DNA-binding proteins targeted at specific loci. Zinc finger proteins (ZFPs), transcription activator-like effectors (TALEs) and nuclease-deactivated Cas9 (dCas9) have thus been employed successfully as tools for direct control of transcription, and – combined with a domain containing nuclease activity – for genome editing (Gaj et al., 2013). These DNA-binding proteins have also been used as vehicles to target epigenetic modifiers to specific loci. This application has been a major step forward in editing the epigenome (de Groote et al., 2012; Thakore et al., 2016). Below, we describe the three adaptable DNA-binding vehicles currently available for targeted recruitment of effectors at designated genomic loci.

ZFPs
Zinc finger proteins (ZFPs) form a large class of DNA-binding proteins that use coordinated zinc ions to stabilize the typical ββα fold in their modular DNA-recognition domains. Each zinc finger domain consists of about 30 amino acids, capable of sequence-specific recognition of a 3–4 base pair sequence (Pavletich and Pabo, 1991). The canonical zinc finger protein harbours three such domains. For increased target specificity, synthetic zinc finger proteins generally consist of 4–6 zinc finger domains arranged in tandem. Choo et al. (1994) demonstrated in 1994 that it is possible to specifically target an oncogenic gene and reduce its transcription through transcriptional blockage by ZFP. This is the first instance of the use of a designed protein for manipulating gene expression at a defined locus. By targeting the histone methyltransferase (HMT) catalytic core using synthetic ZFPs to an endogenous genomic reporter system the causality of histone modifications in inducing repression of transcription was established (Snowden et al., 2002). Histone 3 Lysine 9 methylation (H3K9me) was found to become enriched throughout a 500–1000 bp region around the target site. This spreading of the H3K9me mark, attributed to HP1, a protein associated with heterochromatin and mediator of gene silencing, resulted in repression of transcription (Snowden et al., 2002). These applications of synthetic ZFPs are key examples of an adaptable DNA-binding platform tool, unique at the time, and laid the basis for the use of ZFP fusions in a wide range of applications (Klug, 2010; Urnov et al., 2010; Gersbach et al., 2014).

TALEs
Transcription activator-like effectors (TALEs) are DNA-binding proteins originating from plant pathogenic bacteria that consist of repeated motifs of 33 or 34 amino acids, with residues 12 and 13 – so-called repeat variable di-residues (RVDs) – in each otherwise conserved repeat recognizing one particular base in double stranded DNA (Boch et al., 2009; Moscou and Bogdanove, 2009; Deng et al., 2012). TALE repeats can – by analogy with ZFPs – be designed in tandem to recognize
any DNA sequence of interest, with each repeat recognizing one base via its RVD (Zhang et al., 2011; Mussolino and Cathomen, 2012). Since di-residues specific for each nucleotide have been identified, the rational design of DNA recognition is straightforward (Zhang et al., 2011; Reyon et al., 2012). Nevertheless, target specificity of designed TALEs generally needs to be verified in vitro or in vivo (Morbitzer et al., 2010; Zhang et al., 2011; Grau et al., 2013; Guilinger et al., 2014). However, also TALEs and TALE–effector fusions verified to be target specific and active in a reporter assay may exhibit different activities when targeted at chromosomal sites, attributed to altered chromatin accessibility (Zhang et al., 2011). Although quite recently discovered, TALEs have proven their value next to ZFPs in gene expression modulation and genome editing (Sanjana et al., 2012; Gaj et al., 2013). TALE fusions have also been used to target epigenetic modifiers to genomic loci. TALEs fused to the synthetic transactivation domain VP64 have been shown to up-regulate transcription of the endogenous pluripotency genes SOX2 and KLF4 in human 293FT cells, whereas TALE-VP64 targeted c-MYC and OCT4 genes were unaffected (Zhang et al., 2011). A TALE–TET1 fusion has been shown to decrease methylation of specific CpG dinucleotides in the targeted promoter regions, resulting in increased mRNA expression levels of the targeted genes (Maeder et al., 2013a).

The use of both ZFPs and TALEs has been instrumental in achieving systematic and controlled targeting of effectors to defined genomic loci. However, there are some drawbacks to the use of these DNA-binding proteins. The design of ZFPs to target specific DNA sequences is not as straightforward and modular as suggested above, as side chain–side chain interactions within and between adjacent zinc fingers complicate DNA recognition (Wolfe et al., 2000, 2001), precluding a robust ‘recognition code’, permitting reliable engineering of highly specific DNA-recognition proteins. Although the DNA recognition via defined RVDs is more modular compared to ZFPs, TALE arrays also suffer from off-target binding (Guilinger et al., 2014; Rogers et al., 2015). In fact, it has been shown that the protein context of a TALE repeat influences the DNA-binding specificity of the array (Rogers et al., 2015). A related drawback is that DNA binding of TALEs (Bultmann et al., 2012) and ZFPs (Choo, 1998; Daniel et al., 2002; Vandevenne et al., 2013) is affected by the presence of methylated cytosines in target sequences, making application in genomic contexts more complicated and less flexible. Independent of these drawbacks limiting application, it is important to realize that both tools require re-design of the protein sequence and validation for individual constructs targeting distinct DNA sequences. This process makes adaptation uncertain as well as time and resource consuming. The capability of employing these proteins in multiplex or high-throughput screening applications is limited by this inflexibility of target adaptation.

dCas9

The CRISPR-associated protein 9 (Cas9) is an antiviral enzyme of the Type II clustered regularly interspaced short palindromic repeat (CRISPR) adaptive immune system in prokaryotes (Barrangou et al., 2007; Bhaya et al., 2011). The endogenous enzyme introduces double-strand breaks in DNA using two catalytic domains (RuvC and HNH). Two RNA molecules – a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) – are used to guide Cas9 to its target sequence. A protospacer adjacent motif (PAM) flanking the target sequence recognized by the crRNA acts as an additional determinant in target recognition (Jinek et al., 2012). In biotechnological applications the protein is generally guided by a chimeric single-guide RNA (sgRNA), a hybrid of crRNA and tracrRNA (Fig. 2.1).

Cas9 has been very rapidly and widely adopted (Doudna and Charpentier, 2014) as tool for genome editing (Cho et al., 2013a; Cong et al., 2013; Hwang et al., 2013; Jinek et al., 2013; Mali et al., 2013b). The nuclease-deactivated variant, dCas9 (Fig. 2.1), initially created to establish which catalytic amino acids are necessary for dsDNA cleavage (Jinek et al., 2012), laid the foundation for a whole range of new applications needing site-specific targeting. The dCas9 protein has been adopted as DNA-binding platform for applications as diverse as transcriptional blockage (Bikard et al., 2013; Qi et al., 2013), gene expression modulation (Cheng et al., 2013; Gilbert et al., 2013), epigenetic editing (Hilton et al., 2015; Kearns et al., 2015) and staining of chromosomal regions for live cell imaging (Chen et al., 2013; Anton et al., 2014; Ma et al., 2015). In the next sections we discuss in detail
design strategies and applications of dCas9 for gene expression modulation and epigenetic editing.

**dCas9 design strategies**

Different strategies of varying complexity and efficacy for the use of dCas9 to manipulate the epigenome in a controlled manner have been developed. In essence, all these strategies belong to one of three categories discussed below (Fig. 2.2). Notably, some strategies discussed below have only been used for Cas9, but are equally suitable for use with dCas9.

**Class I – direct effector fusion**

In the most straightforward design, the dCas9 protein (with no effector fused) is used to interfere with transcription via steric blockage of RNA polymerase binding or transcription elongation (Bikard et al., 2013; Qi et al., 2013). This strategy is successful in prokaryotes leading up to ~300-fold mRNA reduction when a single sgRNA is used to target dCas9 or even ~1000-fold when two sgRNAs are combined to block transcription elongation (Qi et al., 2013). However, interference with transcription has not been successful in mammalian cells yielding only ~2-fold reduction in transcript levels (Qi et al., 2013). For potent modulation of gene expression in mammalian cells by dCas9, specific effectors are implied. Down-regulation (CRISPRi) or activation (CRISPRa) of targeted genes is achieved by genetically fusing effector proteins – or their active domains – to dCas9 and expressing them as a single recombinant protein (Fig. 2.2A). Transcription activator domains (VP64, p65) or repressor domains (KRAB, SID) have been fused to dCas9 to specifically increase or decrease target gene expression (Gilbert et al., 2013; Maeder et al., 2013b; Perez-Pinera et al., 2013; Lawhorn et al., 2014). Single dCas9–effector fusions are commonly targeted to adjacent sites using multiple different sgRNAs for maximum impact. In an effort to obtain maximum activation a combination of three different effectors (VP64, p65 and Rta) has been fused in succession to dCas9, resulting in a ~100-fold increase in transactivation of the target genes compared to dCas9-VP64 alone (Chavez et al., 2015). A basic strategy towards achieving temporal control of dCas9–effector target binding is inducible expression of dCas9 or the sgRNA, e.g. by using a doxycycline (González et al., 2014; Wang et al., 2014; Dow et al., 2015) or IPTG-responsive promoter. However, this approach suffers from leaky expression in the absence of an inducer.

**Class II – indirect effector recruitment**

Class II strategies incorporate into the basic design additional motifs that recruit effector proteins (Fig. 2.2B). An example of such a motif is the SunTag (Tanenbaum et al., 2014), a protein scaffold containing peptide epitopes (a variant with 10 and one with 24 epitopes is available) able to recruit effector domains via specific single-chain variable fragment (scFv) antibodies (Fig. 2.2B). The SunTag-carrying
Figure 2.2 Design strategies of using dCas9 to target effector domains to specific DNA sequences. (A) The effector domain is directly fused to dCas9 to recruit it to sequences specified by the sgRNA. (B) The effector domain(s) is recruited via functional scaffolds incorporated in the sgRNA–dCas9 complex, either via fusion to dCas9 (left) or via RNA aptamers in a scaffolding RNA (scRNA, right). (C) Spatiotemporal control of effector activity is obtained via controlled recruitment of effectors to the sgRNA–dCas9 complex (left) or the reconstitution of split-dCas9 directly fused to effectors (right) via light- or chemical-inducible heterodimerization partners.
dCas9 has been shown to successfully recruit scFv–VP64 fusions and to increase target gene expression more than the simple dCas9–VP64 fusion (Tanenbaum et al., 2014).

More complex systems with added functionality are not limited to direct genetic fusion of effector domains to dCas9. The sgRNA can be extended to include RNA aptamers (Fig. 2.2B), secondary RNA structures specifically recognized by RNA-binding proteins, to form a scaffolding RNA (scRNA) (Mali et al., 2013a; Konermann et al., 2015; Zalatan et al., 2015). Using scRNAs with RNA aptamers such as MS2, PP7, com or the PUF binding site (PBS), effectors can be recruited to the dCas9–sgRNA complex indirectly via fusion to corresponding RNA-binding proteins (Mali et al., 2013a; Zalatan et al., 2015; Cheng et al., 2016). Recruiting effectors simultaneously via a dCas9 gene fusion and via aptamers present in the sgRNA has been shown to yield strong synergistic transactivation (Konermann et al., 2015; Xu et al., 2016). For example, recruiting two MS2–p65-HSF1 fusions to dCas9-VP64 via an scRNA containing two MS2 hairpin aptamers resulted in a 100-fold enhancement of transactivation compared to just dCas9-VP64. Using just one scRNA this system outperformed a pool of eight sgRNAs targeting dCas9-VP64 to distinct sites along the proximal promoter region of target genes (Konermann et al., 2015).

**Class III – spatiotemporal control of activity**

In a third class, dCas9 strategies are aimed at precisely controlling effector recruitment or dCas9 DNA-binding activity in space and/or time (Fig. 2.2C). Such control has been achieved by using split-(d)Cas9 or split-(d)Cas9-effector proteins that are conditionally assembled into a functional DNA-binding complex in the presence of sgRNA (Wright et al., 2015), upon chemical induction (Zetsche et al., 2015b) or light induction (Nihongaki et al., 2015). Split-(d)Cas9 reconstitution, the heterodimerization partners CRY2 and CIB1 have been used to bring together full-length dCas9 and effectors such as VP64 or p65 upon blue light irradiation (Konermann et al., 2013; Nihongaki et al., 2015b; Polstein and Gersbach, 2015) (Fig. 2.2C). Although the light-inducible CRY2–CIB1 pair works well to bring together dCas9 and effectors, the use of these partners was unsuccessful when applied to reassembly of split-(d)Cas9 (Nihongaki et al., 2015a). An alternative strategy to control the binding of Cas9 to its target sequence relies on intein-mediated splicing (Davis et al., 2015; Truong et al., 2015). Designs with intein in the reading frame of full-length Cas9 or fused to both fragments of split-Cas9 have been used. In both cases, after intein trans-splicing is induced full-length functional Cas9 is obtained (Davis et al., 2015; Truong et al., 2015). An other, less common approach, is to use Cas9 inactive due to caging of lysine residues necessary for Cas9 function (Hemphill et al., 2015). Exposure to UV light removes the caging group (Riggsbee and Deiters, 2010) and recovers an active Cas9 (Hemphill et al., 2015).

Due to the availability of Cas9 orthologues (Chylinski et al., 2014) and redesigned synthetic Cas9 proteins (Kleinstiver et al., 2015; Hirano et al., 2016) with different PAM recognition sequences (Table 2.1), combinations of multiple dCas9 orthologues and matching sgRNAs can be used in parallel to perform distinct activities. In *E. coli* two orthologues have been used to simultaneously cleave viral dsDNA and repress a reporter gene via transcriptional blockade (using a nuclease-active SpyCas9 and a deactivated NmCas9 protein respectively) (Esvelt et al., 2013). In this study it was established that *N. meningitidis*, *S. thermophilus* CRISPR1 and *S. pyogenes* Cas9 function fully orthogonal to one another in *E. coli* as well as in
human cells, allowing for targeting distinct and non-overlapping sets of sequences within the same cell (Esvelt et al., 2013). The three catalytically deactivated Cas9 orthologues from the same three species have been used to label distinct chromosomal loci in live human cells (Ma et al., 2015).

Applications of the dCas9 tool
Targetable DNA-binding proteins with additional activity in epigenetic editing and/or gene expression modulation can be divided in two categories based on the nature of the effector: (i) targeted transcription factors (TTFs) and (ii) targeted epigenetic modifiers (TEMs) (Table 2.2). Effectors used in TTFs are (derivatives of) natural abundant and potent transcription factors that affect gene expression, but also often have indirect and multilateral effects on the states of the epigenome through recruitment of (multiple) effector partners. Commonly used effectors are herpes simplex viral protein 16, VP16 (Wang et al., 2000; Memedula and Belmont, 2003) – better known in tetrameric form as VP64 (Beerli et al., 1998) – and p65, a subunit of the human NF-κB transcription factor (Schmitz and Baueuerle, 1991). Both effectors exhibit complex mechanisms of transactivation via recruitment of secondary transcription factors (Mittler et al., 2003;
van Essen et al., 2009). As transcription repressor the Krüppel-associated box (KRAB), a domain found in numerous mammalian repressors, is commonly used. Together with recruited co-repressor KAP1, KRAB attracts a variety of epigenetic modifiers and chromatin remodelling proteins to induce heterochromatin and block transcription (Groner et al., 2010). TEMs are designed based

<table>
<thead>
<tr>
<th>Targeted transcription factors (TTFs)</th>
<th>Indirect recruitment</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct fusion</td>
<td>scRNA</td>
<td>SunTag</td>
</tr>
<tr>
<td>Activating</td>
<td>VP64</td>
<td>Chavez et al., 2015; Farzadfard et al., 2013; Gao et al., 2014; Gilbert et al., 2013; Hilton et al., 2015; Hu et al., 2014; Kearns et al., 2014; Maeder et al., 2013b; Mali et al., 2013b; Perez-Pinera et al., 2013; Polstein and Gersbach, 2015; Zetsche et al., 2015a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repressing</td>
<td>KRAB</td>
<td>Farzadfard et al., 2013; Gao et al., 2014; Gilbert et al., 2013; Hu et al., 2014; Kearns et al., 2014, 2015; Lawhorn et al., 2014; Thakore et al., 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targeted epigenetic modifiers (TEMs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histone acetylation</td>
<td>p300</td>
<td>Hilton et al., 2015</td>
</tr>
<tr>
<td>CBP</td>
<td>CBP</td>
<td>Cheng et al., 2016</td>
</tr>
<tr>
<td>Histone demethylation</td>
<td>LSD1</td>
<td>Kearns et al., 2015</td>
</tr>
<tr>
<td>DNA methylation</td>
<td>DNMT3A</td>
<td>McDonald et al., 2016; Vojta et al., 2016</td>
</tr>
<tr>
<td>DNA demethylation</td>
<td>TET1</td>
<td>Xu et al., 2016</td>
</tr>
</tbody>
</table>
Manipulating the epigenome and gene expression using TEMs and TTFs

So far only a few studies have been published reporting the use of dCas9 fusions for epigenome editing and manipulating gene expression. The catalytic histone acetyltransferase (HAT) core domain p300 has been used to catalyse acetylation of histones in human HEK293T cells (Hilton et al., 2015). Targeting dCas9–p300 fusions to promoter regions or proximal or distant enhancers caused activation of gene expression. Increased expression upon enhancer-targeting was concomitant with enrichment in H3K27ac at the corresponding genomic target sites (Hilton et al., 2015). In most cases the same genes could be transactivated by dCas9-VP64 when targeted at promoters. To achieve transactivation both effectors can thus be used. The two effectors behave somewhat differently in terms of their impact on histone acetylation state, as p300 directly catalyses H3K27ac (Ogryzko et al., 1996; Delvecchio et al., 2013), whereas VP64 recruits subsequent transactivation components, amongst which is p300 (Menedula and Belmont, 2003). Also the histone acetyltransferase domain of the CREB-binding protein has been fused to dCas9 (dCas9-CBPHAT) and has been used to catalyse locus-specific acetylation of histones (Cheng et al., 2016). dCas9-CBPHAT was targeted using the Casilio (CRISPR/Cas9-Pumilio) system, which harbours an scRNA containing multiple PUF binding sites (PBS), to recruit additional CBPHAT domains via fusions with Pumilio/FBF (PUF) RNA-binding domains. Similar to dCas9-p300, targeting dCas9-CBPHAT to promoters or proximal and distal enhancer caused increased expression of the target genes (Cheng et al., 2016).

dCas9 has also been used to introduce DNA methylation by targeting the catalytic domain of the de novo DNA methyltransferase 3A (DNMT3A) to specific loci. dCas9-DNMT3A has been used to methylate cytosines around the transcription start site (TSS) of human and mice genes (McDonald et al., 2016; Vojta et al., 2016). Using this strategy, the methylation status of the CpG island spanning the TSS of the tumour-suppressor gene CDKN2A increased by 20%, resulting in 40% reduction in CDKN2A mRNA levels (McDonald et al., 2016). Targeting the unmethylated promoter regions of the BACH or IL6ST genes increased CpG methylation by 50% and reduced expression 2-fold (Vojta et al., 2016). However, this increase in methylation is not the only cause of reduced expression; catalytically inactive DNMT3A also induced repression, attributed to steric hindrance of the transcription machinery (McDonald et al., 2016; Vojta et al., 2016). In contrast, when dCas9-DNMT3A was targeted to a CpG island 100–400 bp downstream of the TSS, resulting in an increase of 20–35% methylation of CpG residues, no significant change in CDKN2A mRNA expression was observed (McDonald et al., 2016). Upon targeting dCas9-DNMT3A to the unmethylated Cdkn1a gene in mouse myeloid progenitor cells resulting in Cdkn1a repression, proliferation was enhanced. In these studies dCas9 was as effective in introducing DNA methylation as ZFP- or TALE-based DNMT3A systems. The highest efficiency of methylation of targeted CpG residues was obtained when target sites are bracketed by inwardly directed sgRNAs (i.e. target CpGs are downstream of the PAM) and within ~50 bp of the sgRNA binding site (McDonald et al., 2016; Vojta et al., 2016).

Recently the catalytic domain of TET1 was used to induce demethylation of DNA (Xu et al., 2016). In this study the catalytic domain of TET1 was fused to dCas9 and two additional copies of the same effector were directed to the target location via two MS2 aptamers integrated in the sgRNA design (Xu et al., 2016). Transiently delivered scRNA/dCas9-TET1 induced demethylation of the targeted RANKL gene, a gene silenced through hypermethylation in HEK-293FT cells, and about 2-fold increased RANKL mRNA levels. Stable expression of the system yielded stronger effects:
10- to 20-fold transactivation (Xu et al., 2016). To demonstrate the generic applicability of this epigenome editing strategy, expression of MAGEB2 and MMP2, which are both silenced by hypermethylation in HeLa cells, was shown to be up-regulated by targeted demethylation. Interestingly, combining multiple sgRNAs that target distinct sites in the RANKL and MAGE genes does not result in significant additive or synergistic increase in transactivation (Xu et al., 2016). A similar non-additive effect in transactivation was observed for dCas9-p300 targeted to multiple adjacent sites to induce histone acetylation (Hilton et al., 2015).

Characterizing known or discovering new regulatory elements in the genome

Large-scale mapping of the human epigenome and chromatin states reveals vast numbers of putative regulatory elements (ENCODE Project Consortium, 2012; Thurman et al., 2012; Roadmaps Epigenomics Consortium et al., 2015). The validation and characterization of these elements benefits from the use of targetable epigenetic effectors to establish the functions of these regions in the genome.

In order to establish whether dCas9-LSD1 can be used to investigate enhancer function, the fusion protein has been targeted at well-characterized and putative enhancers in mouse embryonic stem cells. Before the introduction of dCas9, TALE–LSD1 fusion proteins had been already used to target endogenous candidate regulatory elements enriched for H3K4me2 and H3K27ac (Mendenhall et al., 2013). These studies established that at a subset of putative enhancers the epigenetic state was altered following LSD1 targeting, permitting identification of the down-regulated target genes (Mendenhall et al., 2013). Targeting of LSD1 to eight putative pluripotency-specific enhancers using dCas9 resulted in the discovery of four enhancers regulating the expression of genes critical in maintaining embryonic state (Kearns et al., 2015). Chromosome conformation capture (3C) confirmed that the target gene was not down-regulated through enhancer–promoter loop disturbance of dCas9-LSD1 binding per se. A reduction in H3K4me2 and H3K27ac was observed around the sgRNA-dCas9-LSD1 target site in the identified Tbx3-enhancer. In contrast, dCas9-KRAB targeted to the same site in the Tbx3-enhancer resulted in a reduction of H3K27ac and an increase of H3K27me3 and H3K9me3 at the Tbx3 proximal promoter region. However, neither dCas9-KRAB nor dCas9-LSD1 was able to increase the level of the repressive marks H3K27me3 and H3K9me3 at the Tbx3-enhancer region. In a different study, in which the KRAB domain was targeted at the HS2 enhancer in the globin locus control region (LCR) in human K562 cells, H3K9me3 marks were effectively introduced at the target enhancer region (Thakore et al., 2015). The introduction of these repressive marks was concurrent with decreased chromatin accessibility at the target site in the enhancer as well as in several other parts of the LCR (Thakore et al., 2015). These findings suggest pleiotropic effects of dCas9-KRAB at the targeted enhancer–promoter loop or as yet uncharacterized downstream mechanisms of chromatin reorganization.

Genome-wide forward screening

Using libraries of targeted DNA-binding platforms fused to effectors the epigenetic states and/or expression levels of genes can be altered to study their impact on selected downstream processes or phenotypes in genome-wide forward screens. Such screening applications gain momentum due to expanding number of high-resolution techniques to measure epigenetic landscapes and chromatin states (Zentner and Henikoff, 2014; Ramani et al., 2016), including the introduction of novel methods for mapping epigenetic modifications in the genomes of single cells (Angermueller et al., 2016; Song et al., 2016).

The application of dCas9 in forward screening using targeted transcription factors (TTFs) was shown to permit highly specific genome-scale transcription modulation (Gilbert et al., 2014; Konermann et al., 2015). Using combinations of 10 sgRNAs per gene, tiling −50 to +300 bp for repression (CRISPRi) or −400 to −50 bp for activation (CRISPRa) around the TSS to target nearly 1600 protein-encoding genes, human K562 cells were screened for growth phenotypes. In this study transcription was repressed using dCas9-KRAB and activated using dCas9-SunTag recruiting scFv–VP64 fusions (Gilbert et al., 2014). In a second screen, known as well as novel complexes and pathways involved in the response to a chimeric
cholera–diphtheria toxin were identified (Gilbert et al., 2014). Using next-generation sequencing, sgRNAs guiding the dCas9-effectors could afterwards be identified to determine the genomic locus responsible for the screened phenotypic characteristics (Gilbert et al., 2014; Konermann et al., 2015). A crucial aspect of screening studies using dCas9–effector fusions is that due to improved sgRNA design and more effective dCas9 design strategies, one dCas9 recruited by a single sgRNA now suffices in modulating a specific locus (Konermann et al., 2015).

Current hurdles in application of dCas9

dCas9 has been shown to exhibit off-target binding at genome-wide scale (Kuscu et al., 2014; Tsai et al., 2015). Studies on target specificity have mainly been done on Cas9 (Fu et al., 2014; Slaymaker et al., 2016; Kleinstiver et al., 2016); these findings can be extrapolated to dCas9 and dCas9-based effectors. The ability of the Cas9–sgRNA complex to localize and bind at a target sequence could be altered by genome accessibility as a consequence of local chromatin environment (Knight et al., 2015) or binding of endogenous proteins such as transcription factors (Hilton et al., 2015). Taking that notion into account target sites can be selected based on information regarding genome-bound proteins (e.g. using DNase-seq) (Maeder et al., 2013b; Kearns et al., 2014; Thakore et al., 2015). Alternatively, DNA accessibility and off-target binding are determined ad hoc for every application. Off-target binding is related to flexibility in PAM sequence recognition and permitted mismatches in the PAM-distal part of the sgRNA (Kim et al., 2015; Ran et al., 2015; Tsai et al., 2015; Wang et al., 2015b; Leenay et al., 2016). Efforts to reduce off-target binding include rational redesign of Cas9 for higher PAM specificity (Kleinstiver et al., 2015), reduction of non-specific charge interactions between Cas9 and non-target DNA strand (Slaymaker et al., 2016) or hydrogen bonding between Cas9 and the backbones of the target DNA strand and the sgRNA (Kleinstiver et al., 2016), as well as use of truncated sgRNAs (Fu et al., 2014). Although many of these strategies lead to a reduction of the DNA binding affinity of Cas9 or its activity, these efforts have contributed significantly to overcoming Cas9 off-target binding.

Another issue for applications of dCas9 is its size. The coding sequence of the widely used SpyCas9 is around 4.2 kb, which is compatible with the ~4.5 kb maximum accommodated in common viral vectors used for delivery in mammalian cells. However, such a long dCas9 coding sequence does not leave ‘space’ for sgRNA sequences and regulatory elements to be co-delivered. This limits the application of some more complex design strategies such as extending the sgRNA with aptamers, adding effector domains to dCas9 or using pools of multiple sgRNAs. The relatively small Cas9 orthologue from Staphylococcus aureus (SaCas9) is 3.2 kb in size (Friedland et al., 2015; Ran et al., 2015). The use of such small orthologues reduces the size problem with no drawbacks on activity, as observed for a truncated SpyCas9 derivative in which a less-conserved, non-interacting portion of the recognition lobe had been removed (Nishimasu et al., 2014). The split-dCas9 system also provides a solution to the size problem, but it requires two vectors for delivery (Truong et al., 2015). Alternatively, cells stably expressing dCas9-effector proteins can be complemented by delivery of sgRNA, but stable expression is difficult to achieve when using primary cells or multicellular organisms. Another option is dCas9 delivery using in vitro assembled sgRNA-(d)Cas9 ribonucleoproteins (RNPs), an approach which was successful for dCas9-VP64 (Zuris et al., 2015) and Cas9 genome editing applications (Cho et al., 2013b; Kim et al., 2014; Ramakrishna et al., 2014; Sun et al., 2015; Zuris et al., 2015; D’Aстолфо et al., 2015). Finally, a potential hurdle for implementation in living organisms is the fact that the (d)Cas9 protein elicits an immunogenic responses due to its exogenous origin (Wang et al., 2015a). Future dedicated studies will need to clarify the extent of this issue for possible clinical applications.

Conclusions and future perspectives

The importance of the epigenome is becoming increasingly clear in the light of developmental processes, cell differentiation status and diseases. In an attempt to draw causal relations rather than correlations to understand and control the interplay between the epigenome, chromatin structure and gene expression levels, targetable DNA-binding proteins have shown to be of great value. dCas9 has a more robust yet readily adjustable DNA-recognition
mechanism compared to ZFPs and TALEs; it has proven to be a flexible tool in recruiting effector domains to targeted DNA-sequences, in particular due to the numerous options arising from expanding the sgRNA even up to 5 kb long autonomous functional RNAs (Shechner et al., 2015). The full value of dCas9 in the field of epigenetics has yet to be revealed with more functional studies at local scales and genome-wide level, providing detailed causal relations between epigenetic modifications, chromatin structure and gene expression. Future application in translational sciences such as medicine, synthetic biology or biotechnology, rely on advances in overcoming final hurdles provided by delivery problems due to its size, possible immunogenic response and off-target effects.

Alternatives for dCas9 emerge from the discovery and characterization of novel and known CRISPR systems involving protein complexes with different DNA-recognition mechanisms or different dimensions. Possible alternatives have recently become available in the form of the relatively small Cpf1 (~3.9 kb) (Zetsche et al., 2015a; Yamano et al., 2016), the C2c1 and C2c3 systems (Shmakov et al., 2015) or even the multiprotein Cascade complex from the abundant Type I-E system native to E. coli K-12 (Luo et al., 2015). These and future related systems add to a comprehensive CRISPR-based toolbox with distinct and orthogonal guiding crRNA, protospacer- or PAM-recognition, which is destined to occupy a permanent and prominent place in molecular biology.

Glossary
VP64: A strong transcriptional activation domain that recruits a variety of transcription factors and chromatin remodelling factors (Cress and Triezenberg, 1991). Tetrameric fusion of the acidic domain of herpes simplex viral protein 16 (VP16) (Beerli et al., 1998).
p65: The 65 kDa subunit of the NF-xB transcription factor (Schmitz and Baeuerle, 1991). Often the principal transactivation domain (TA1) is used as activation domain (p65AD).
KRAB: Krüppel-associated box (KRAB) is a common repression domain in eukaryotic transcription factors (Margolin et al., 1994). It associates with KAP1, which forms a scaffold to recruit several proteins involved in inducing and spreading of heterochromatin over large distances (Groner et al., 2010).
p300: Human transcriptional regulatory protein able to acetylate histone 3 at lysine residue 27 (Ogryzko et al., 1996), a signal for active transcription. Often the isolated catalytic core of p300, containing the histone acetyltransferase (HAT) domain, is used as effector.
CBP: The CREB-binding protein (CBP) is able to bind many different transcription factors, acting as scaffold to co-activate transcription (Chan and La Thangue, 2001). Additionally, it has a histone acetyl transferase (HAT) domain, which can be used as epigenetic effector to acetylate lysine residues in histones.
LSD1: Lysine-specific demethylase 1 (LSD1) demethylates histone 3 at lysine residues 4 and 9, leading to silencing of enhancers (Shi et al., 2004). LSD1 functionally interacts with other chromatin-modifying enzymes, including histone deacetylases that remove acetyl groups from histone 3 at lysine residue 27 (Lee et al., 2006).
DNMT3A: DNA methyltransferase 3A (DNMT3A) catalyses de novo DNA methylation as well as methylation of hemimethylated DNA, preferentially at CpG sites (Okano et al., 1998; Oka et al., 2006). Methylated CpGs in promoters are associated with silenced transcription.
TET1: Ten-eleven translocation (TET) methylcytosine dioxygenase (TET1) is able to oxidate 5-methyl cytosines (5mC) to 5-hydroxymethylcytosines (5hmC), a step in the active removal of 5mC (Tahiliani et al., 2009; Xu et al., 2011). At promoters, this oxidation and removal is associated with activation of silenced genes.

Acknowledgements
Research on the topic of this review in the lab of R.T.D. is supported by a grant from the Netherlands Organization for Scientific Research (VICI 016.160.613), M.T.-D. is supported by a grant from the Dutch Technology Foundation STW. D.J.W.B. contributed to the writing of this review article as part of the MSc programme in Life Science and Technology at Leiden University.

References


Brocken et al.


