Research

Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators

Lauren R. Polstein,[1] Pablo Perez-Pinera,[1,7] D. Dewran Kocak,[1] Christopher M. Vockley,[2,3] Peggy Bledsoe,[2] Lingyun Song,[2] Alexias Safi,[2] Gregory E. Crawford,[2,4] Timothy E. Reddy,[2,5] and Charles A. Gersbach[1,2,6]

[1]Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, USA; [2]Center for Genomic and Computational Biology, Duke University, Durham, North Carolina 27708, USA; [3]Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA; [4]Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, North Carolina 27710, USA; [5]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina 27710, USA; [6]Department of Orthopaedic Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA

Genome engineering technologies based on the CRISPR/Cas9 and TALE systems are enabling new approaches in science and biotechnology. However, the specificity of these tools in complex genomes and the role of chromatin structure in determining DNA binding are not well understood. We analyzed the genome-wide effects of TALE- and CRISPR-based transcriptional activators in human cells using ChIP-seq to assess DNA-binding specificity and RNA-seq to measure the specificity of perturbing the transcriptome. Additionally, DNase-seq was used to assess genome-wide chromatin remodeling that occurs as a result of their action. Our results show that these transcription factors are highly specific in both DNA binding and gene regulation and are able to open targeted regions of closed chromatin independent of gene activation. Collectively, these results underscore the potential for these technologies to make precise changes to gene expression for gene and cell therapies or fundamental studies of gene function.

[Supplemental material is available for this article.]

Recently developed genome engineering technologies are powering new advances and approaches in genomics, genetics, and gene therapy (Gaj et al. 2013; Gersbach and Perez-Pinera 2014). These tools include approaches for editing genome sequences using site-specific nucleases and controlling gene expression with targeted activators, repressors, or other modifiers of the epigenome. Although these methods are already being applied in diverse contexts, important questions remain about the specificity of their action in complex genomes and how they access target sites in various chromatin states.

The discovery of the modular DNA recognition code of transcription activator-like effectors (TALEs) (Boch et al. 2009; Moscou and Bogdanove 2009), DNA-binding proteins that exist in plant-pathogenic bacteria, led to the creation of robust engineering tools that precisely modify cellular genomes. TALE proteins targeted to new DNA sequences can be easily generated through the assembly of domains that recognize each of the four nucleotides (Bogdanove and Voytas 2011; Cermak et al. 2011). These DNA-binding proteins can then be fused to nuclease domains (Christian et al.

2010; Miller et al. 2011) or regulatory and epigenome-modifying domains (Zhang et al. 2011; Cong et al. 2012; Konermann et al. 2013; Maeder et al. 2013a; Mendenhall et al. 2013) to achieve targeted genome engineering.

More recently, the clustered regularly interspaced short repeats (CRISPR)-associated (Cas) system has emerged as an extremely powerful and facile technology for genome engineering (Hsu et al. 2014; Sander and Joung 2014). The engineered CRISPR system, which has been repurposed from a naturally occurring mechanism of bacterial adaptive immunity (Wiedenheft et al. 2012), consists of the Cas9 nuclease and a short guide RNA (gRNA) that forms a complex with Cas9 and directs it to a 20-bp target sequence in the genome through complementary base pair hybridization (Jinek et al. 2012). The only sequence restriction of the 20-bp target site, known as the protospacer, is that it must be immediately adjacent to a short sequence referred to as the protospacer-adjacent motif (PAM). For example, the natural PAM sequence for the Cas9 from *Streptococcus pyogenes*, the most commonly used CRISPR system, is 5′-NGG-3′. This CRISPR system can be used in orthogonal species for genome editing (Cho et al. 2013; Cong

1158 **Genome Research**
www.genome.org
25:1158–1169 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/15; www.genome.org

et al. 2013; Hwang et al. 2013; Jinek et al. 2013; Mali et al. 2013b), and a catalytically inactive form of Cas9 (dCas9) can be fused to regulatory domains for targeted gene regulation (Cheng et al. 2013; Farzadfard et al. 2013; Gilbert et al. 2013; Maeder et al. 2013b; Perez-Pinera et al. 2013a; Qi et al. 2013; Hilton et al. 2015; Kearns et al. 2015).

Despite the widespread use of the TALE and CRISPR technologies for diverse applications, there continues to be considerable uncertainty regarding the specificity of their action in the context of complex human genomes. Studies of TALE nuclease (TALEN) specificity in human cells have readily detected activity at off-target sites with sequence homology to the intended target site using both bioinformatic predictions (Hockemeyer et al. 2011; Fine et al. 2014) and approaches for genetically labeling sites of nuclease activity within cells (Osborn et al. 2013); however, clonal populations without modification of the exome can be easily obtained (Ousterout et al. 2013). Other assays of DNA-recognition properties of purified TALEs or TALENs showed that although the proteins are highly specific for the intended target site, there are significant levels of activity at sites containing sequence mismatches (Meckler et al. 2013; Guilinger et al. 2014). Although gene regulation with dCas9 was shown to be exceptionally specific by RNA sequencing (Gilbert et al. 2013; Perez-Pinera et al. 2013a) and microarrays (Cheng et al. 2013), there have been several reports of significant levels of off-target Cas9 nuclease activity in human cells (Cradick et al. 2013; Hsu et al. 2013; Mali et al. 2013a; Pattanayak et al. 2013; Cho et al. 2014; Fu et al. 2014). These studies of CRISPR/Cas9 specificity have used various methods to determine off-target sites, including bioinformatic prediction (Fu et al. 2013; Hsu et al. 2013), high-throughput reporter assays (Mali et al. 2013a), and profiling activity of a purified CRISPR system in vitro (Pattanayak et al. 2013). More, recently, methods for genome-wide interrogation of nuclease activity have been developed (Frock et al. 2015; Kim et al. 2015; Tsai et al. 2015; Wang et al. 2015). However, unbiased genome-wide methods for quantitatively determining target site binding and subsequent gene regulation function of these genome engineering tools have not been as broadly explored.

It is also necessary to better understand which regions of the genome are targetable by these technologies. Beginning with early work with engineered zinc finger transcription factors (Liu et al. 2001), there has been an assumption that only regions of open chromatin can be targeted. Thus, these engineered proteins were only targeted to DNase I hypersensitive regions, and this approach has continued in more recent work with TALE- and CRISPR-based transcription factors (Maeder et al. 2013b,c). In contrast, we recently demonstrated the activation of silent genes by targeting promoters in heterochromatin with both of these technologies (Perez-Pinera et al. 2013a,b). However, the effects of these proteins on local and genome-wide chromatin state following gene activation remain poorly understood.

Determining the specificity and impact of chromatin state on the function of these modern genome engineering tools is critical to their further development and to the interpretation of results obtained with these systems. In order to address these concerns, we performed a genome-wide analysis of DNA binding, gene regulation, and chromatin remodeling using ChIP-seq, RNA-seq, and DNase-seq, respectively, for both TALE- and CRISPR/dCas9-based transcription factors. Our results show exceptional genome-wide specificity for both technologies in human cells by all three assays, which is unexpected considering previous reports. These observations are significant for interpreting the results of predictive or biased specificity assays, choosing genomic target sites, and designing improved genome engineering tools.

Results

To examine the genome-wide specificity of DNA binding of both TALE- and CRISPR/Cas9-based genome engineering tools, we chose to study transcriptional activators (TALE-VP64 and dCas9-VP64) rather than nucleases in order to decouple target site recognition from DNA cleavage and error-prone DNA repair by nonhomologous end joining that could affect DNA-binding properties. Additionally, this allowed us to focus on TALE monomers rather than the heterodimer necessary for TALEN activity at an endogenous genomic target site. Finally, this permitted the analysis of chromatin remodeling concomitant with the targeted activation of silent genes. As model target genes to study gene activation and corresponding chromatin remodeling, we chose the *IL1RN* and *HBG1/2* loci because (1) they are not expressed in the HEK293T cell line that we used for these studies; (2) their promoters do not contain DNase I hypersensitive (DHS) sites in HEK293T cells (Perez-Pinera et al. 2013a,b); and (3) the products of these genes, the IL-1 receptor antagonist (IL1RA) and gamma globin, do not have any known effect on transcription in these cells, allowing us to examine the primary effects mediated by TALE- and dCas9-based genome engineering tools by RNA-seq and DNase-seq analysis. Furthermore, both genes encode proteins with biomedical relevance, as IL1RA is an approved anti-inflammatory biologic drug (anakinra), and activation of gamma globin expression is a focus of therapies for sickle cell disease.

The TALE- and CRISPR/Cas9-based transcriptional activator technology has most commonly been applied with combinations of these engineered factors targeted to individual promoters in order to generate robust changes in gene expression (Cheng et al. 2013; Farzadfard et al. 2013; Maeder et al. 2013b,c; Mali et al. 2013a; Perez-Pinera et al. 2013a,b; Kabadi and Gersbach 2014). In order to assess if this approach may result in unanticipated off-target effects, we also used combinations of four TALEs or gRNAs targeted to each promoter (Fig. 1A; Supplemental Table 1). This also allowed us to investigate genome-wide specificity under experimental conditions that are known to generate robust changes in gene expression (Perez-Pinera et al. 2013a,b). Because the *HBG2* gene is a duplication of the nearby *HBG1* gene, three of the four TALEs and gRNAs (Supplemental Table 1, A–C) perfectly recognize sites in the *HBG2* promoter as well. Expression plasmids for each TALE-VP64 or each gRNA and dCas9-VP64 were transfected into HEK293T cells, and gene activation was measured by qRT-PCR. Consistent with our previous studies (Perez-Pinera et al. 2013a,b), expression of a single TALE-VP64 or dCas9-VP64 delivered with a single gRNA led to modest effects on gene expression, whereas combinations of TALE-VP64s or gRNAs led to robust gene activation (Fig. 1B; Supplemental Tables 2–5). Activation with TALE-VP64 was substantially greater than activation with dCas9-VP64, consistent with previous reports (Konermann et al. 2013; Maeder et al. 2013b,c; Perez-Pinera et al. 2013a,b; Gao et al. 2014). Importantly, the activation domain was critical to inducing gene expression, as delivery of the TALEs or dCas9 and gRNAs without the VP64 domain did not have any effect on expression levels (Fig. 1C).

We have previously shown by RNA-seq that targeted activation of the *IL1RN* gene with dCas9-VP64 is exceptionally specific, with no other significantly up-regulated genes and only one significant down-regulated gene (Perez-Pinera et al. 2013a). A similar
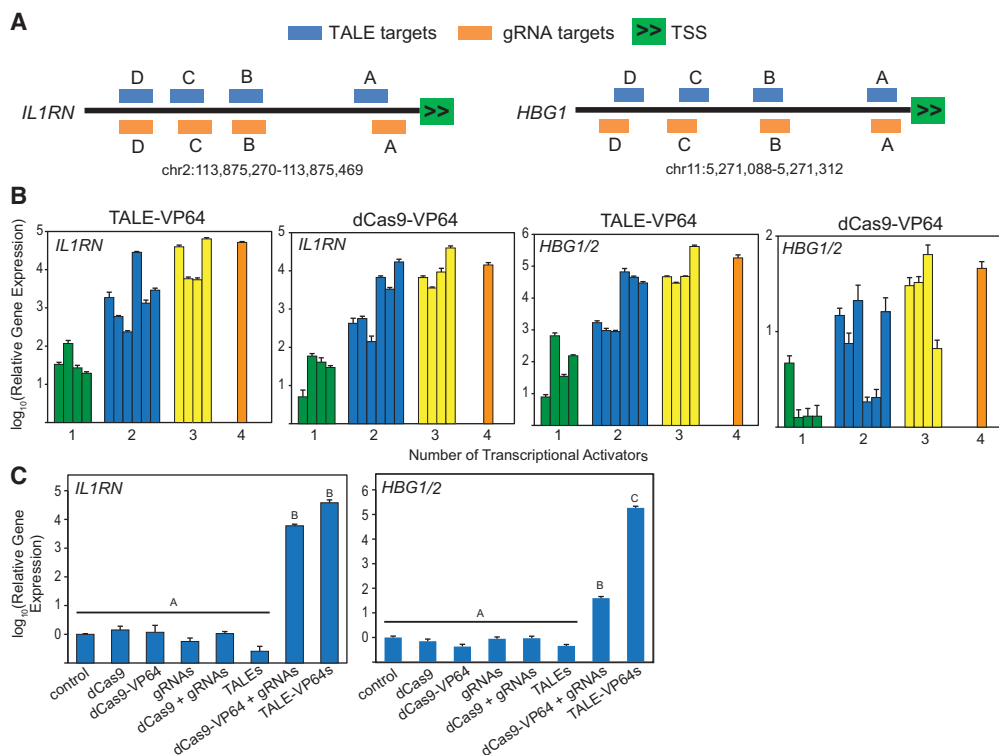
Figure 1. Targeted activation of the human *IL1RN* and *HBG1/2* genes by TALE-VP64 and dCas9-VP64 transcription factors. (*A*) Four TALEs (blue) and four gRNAs (orange), each labeled A through D, were designed to target the *IL1RN* and *HBG1/2* promoters within the ~200 bp upstream of the transcriptional start site (TSS; green). The position of each TALE and gRNA is shown to scale. (*B*) Single expression plasmids or combinations of two, three, or four expression plasmids for the TALE-VP64s or gRNAs, along with dCas9-VP64, targeted to each gene were transfected into HEK293T cells. Expression of the target gene was assessed by qRT-PCR. Robust gene activation was observed only in response to the combination of TALE-VP64s or gRNAs with dCas9-VP64. (*C*) The VP64 activation domain is essential for target gene induction. HEK293T cells were transfected with the combination of expression plasmids for the four TALEs with and without the VP64 domain, or four gRNAs either alone or with dCas9 or dCas9-VP64. Only samples transfected with TALE-VP64s or gRNAs with dCas9-VP64 showed changes in target gene expression. Gene expression is normalized to *GAPDH* levels and shown as fold-increase relative to control cells transfected with an empty expression plasmid (mean ± SEM, *n* = 4 independent transfections across two experiments, different letters indicate *P* < 0.0001 by Tukey's test after log transformation).

level of specificity was achieved at the *HBG1/2* locus, although the overall level of activation was not enough to be statistically significant after multiple hypothesis testing, consistent with the qRT-PCR results showing much weaker activation of *HBG1/2* by dCas9-VP64 compared to *IL1RN* (Fig. 1B). To determine whether activation with TALE-VP64 had a similar level of specificity, we repeated this analysis on HEK293T cells transfected with the combination of four TALE-VP64s targeting the *IL1RN* and the *HBG1/2* promoters or with empty plasmid as a control (Supplemental Tables 6, 7). Activation of the four isoforms of the *IL1RN* gene by TALE-VP64 was again robust (2.5–3.3× over control) and significant ($P = 10^{-4}$–10^{-6}) with a false discovery rate across all genes tested of 3% for the most significant isoform (Fig. 2A). Meanwhile, *HBG1* and *HBG2* increased in expression 65-fold ($P = 10^{-15}$) and 79-fold ($P = 10^{-17}$), respectively, relative to control in response to the TALE-VP64s (Fig. 2C). Differences in fold-activation between qRT-PCR and RNA-seq data are the result of the much lower detection limit for qRT-PCR. For both *IL1RN* and *HBG1/2*, the specificity of gene activation by TALE-VP64 was exceptional, as no other genes were identified as differentially expressed with a false discovery rate <95%. To further investigate the possibility of off-target effects that are general to any TALE-VP64 and do not depend on the particular targeted sequence, we combined replicates of the *IL1RN* and *HBG1/2* experiments and compared them to the con-

trol condition. There were no changes to gene expression that achieved genome-wide significance, although the analysis revealed two genes with nominally significant expression changes ($P \sim 0.01$ before correction for multiple hypothesis testing), including *MUC4* and *HRC* that were induced 1.8-fold and 1.7-fold, respectively. To compare the magnitude of genome-wide gene activation between TALE-VP64 and dCas9-VP64, we plotted the level of expression of every gene in samples treated with dCas9-VP64 and gRNAs versus samples treated with TALE-VP64s for both *IL1RN* (Fig. 2B) and *HBG1/2* (Fig. 2D). Corroborating our qRT-PCR results (Fig. 1B), and previous observations by us and others (Konermann et al. 2013; Maeder et al. 2013b; Mali et al. 2013a; Perez-Pinera et al. 2013a; Gao et al. 2014), the TALE-VP64s led to higher levels of activation for both genes. This analysis also shows the absence of any significant off-target effects that are specific to TALE- or dCas9-based activators (Fig. 2D).

Although the effects of TALE-VP64 and dCas9-VP64 on gene activation were highly specific, this does not exclude binding to off-target sites that do not affect gene expression, either because the binding site is far from any gene or regulatory element, or because only a single activator protein bound to an off-target site was not sufficient to generate significant changes in gene expression that result from synergistic action of multiple activators (Maeder et al. 2013b,c; Perez-Pinera et al. 2013a,b). To determine
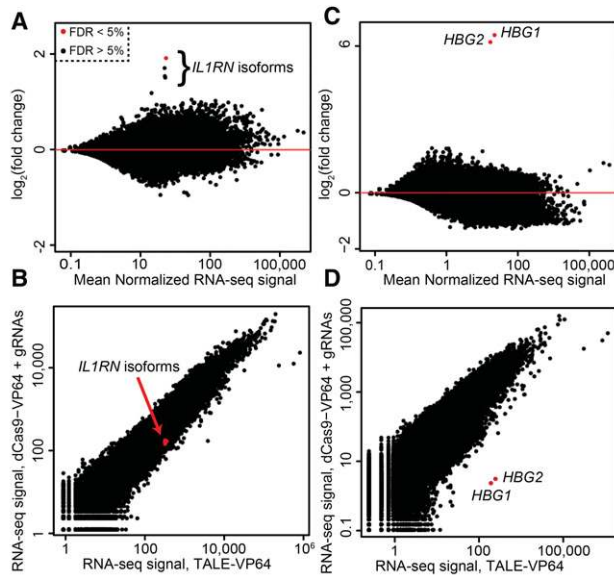
Figure 2. Genome-wide specificity of TALE-VP64 and dCas9-VP64-mediated gene activation. RNA-seq was performed on samples co-transfected with a set of four TALE-VP64 expression plasmids targeting either *IL1RN* (*A*) or *HBG1/2* (*C*). In each case, the only genome-wide significant changes (false discovery rate <5%) in gene expression between the treatments and an empty plasmid-transfected control were increases in the expression of *IL1RN* or *HBG1/2*, respectively. (*B,D*) Comparison of RNA-seq measurements of gene expression after activating expression using TALE-VP64 (*x*-axis) or dCas9-VP64 (*y*-axis). (*B*) When targeting *IL1RN*, the TALE-VP64-mediated activation was slightly stronger. (*D*) When targeting *HBG1/2*, TALE-VP64 had a substantially stronger effect on expression. RNA-seq for dCas9-VP64 samples was published previously (Perez-Pinera et al. 2013a).
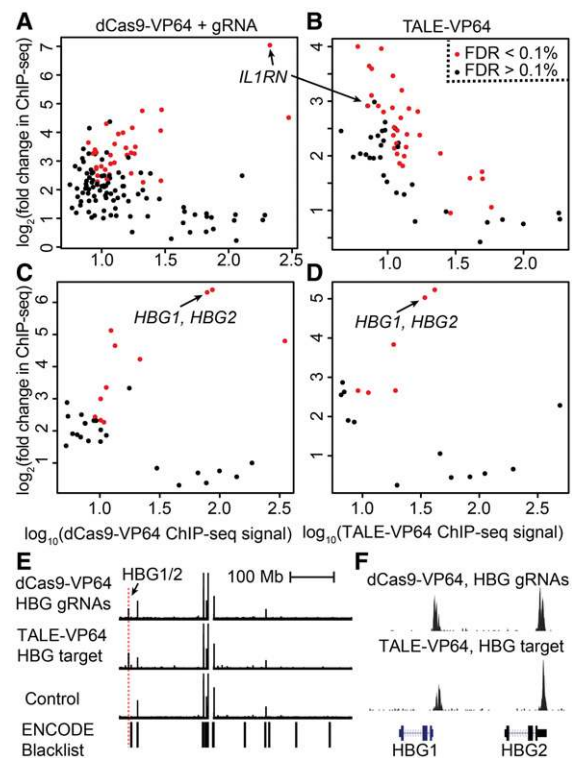
the genome-wide binding specificity of these proteins, we performed ChIP-seq for the HA epitope tag present on TALE-VP64 and dCas9-VP64. Samples included HEK293T cells transfected with combinations of plasmids encoding the four TALE-VP64s or the combination of plasmids encoding the four gRNAs and dCas9-VP64 targeting either the *IL1RN* or *HBG1/2* promoters. Cells transfected with an empty expression plasmid were included as a control, and three biological replicates, transfected on different days, were used for each condition. We only considered binding sites that were reproducible in two of three biological replicates with an irreproducible discovery rate (IDR) <0.05 (Landt et al. 2012) and that had a significant increase in ChIP-seq signal according to a negative-binomial background model that allows for overdispersion in sequencing count reads (Anders and Huber 2010).

For both TALE-VP64 and dCas9-VP64 proteins targeting the *IL1RN* promoter, we identified binding at the target sites as well as 31 off-target binding sites (Fig. 3A,B; Supplemental Tables 8, 9). For dCas9-VP64, the target site in the *IL1RN* promoter region had the strongest evidence of binding (131-fold increase in ChIP-seq signal over control). The off-target sites had notably weaker increases in signal strength (range: from 4.8- to 27.6-fold increase in signal over control) (Supplemental Table 9). For TALE-VP64 targeting the same locus, we observed a 7.3-fold increase in ChIP-seq signal at the target site. Unlike the results for dCas9-VP64, however, the off-target binding sites for the TALE-VP64 protein had an overlapping distribution of increases in signal strength (range: from 1.93- to 16-fold) (Supplemental Table 8).

In the case of TALE-VP64 and dCas9-VP64 targeted to the *HBG1/2* promoters, we obtained similar specificity results (Fig.

3C,D; Supplemental Tables 10, 11). In both cases, we identified the intended target sites and four and nine off-target binding sites for TALE-VP64 and dCas9-VP64, respectively. Of the binding sites identified, the *HBG1/2* target sites had the greatest increase in ChIP-seq signal in each experiment (~84-fold increase in ChIP-seq signal for dCas9-VP64; ~37-fold increase for TALE-VP64). Off-target dCas9-VP64 binding sites had ChIP-seq signal increases between five- and 35-fold, whereas off-target TALE-VP64 binding sites had signal increases in the range of six- to 14-fold. Based on these results, we conclude that both TALE-VP64 and dCas9-VP64 binding was highly specific across the genome (Fig. 3E,F).

To determine the TALE and gRNA/dCas9 recognition sequences within identified binding sites and to investigate mechanisms of off-target binding, we performed de novo motif detection on all target sites identified by ChIP-seq. For each combination of either TALEs or gRNAs targeted to either *IL1RN* or *HBG1/2*,

Figure 3. Genome-wide specificity of dCas9-VP64 and TALE-VP64 DNA-binding. ChIP-seq was used to map the genomic locations of dCas9-VP64 targeted to the *IL1RN* promoter (*A*), TALE-VP64 targeted to the *IL1RN* promoter (*B*), dCas9-VP64 targeted to the *HBG1/2* promoters (*C*), or TALE-VP64 targeted to the *HBG1/2* promoters (*D*). In each plot, points are binding sites that are reproducible in at least two of three replicates. The *x*-axes are the mean ChIP-seq signal, and the *y*-axes are the fold-change in signal in samples transfected with dCas9-VP64 or TALE-VP64 transcription factors compared to controls. Red points represent binding sites with a statistically significant increase in signal strength according to analysis with DESeq (false discovery rate <0.1%). (*E*) As an example of the genome-wide binding specificity dCas9-VP64 and TALE-VP64 targeting, ChIP-seq signal from experiments targeting *HBG1/2* was plotted across Chromosome 11. ChIP-seq signals found in both experimental conditions and in the control condition are also found in the ENCODE blacklist, indicating that they are technical artifacts of ChIP-seq and not binding events. Meanwhile, strong ChIP-seq signal was found at the *HBG1/2* promoters in the dCas9-VP64 and TALE-VP64 conditions but not in the control condition. (*F*) The dCas9-VP64 and TALE-VP64 ChIP-seq peaks localize to the intended *HBG1/2* promoters.
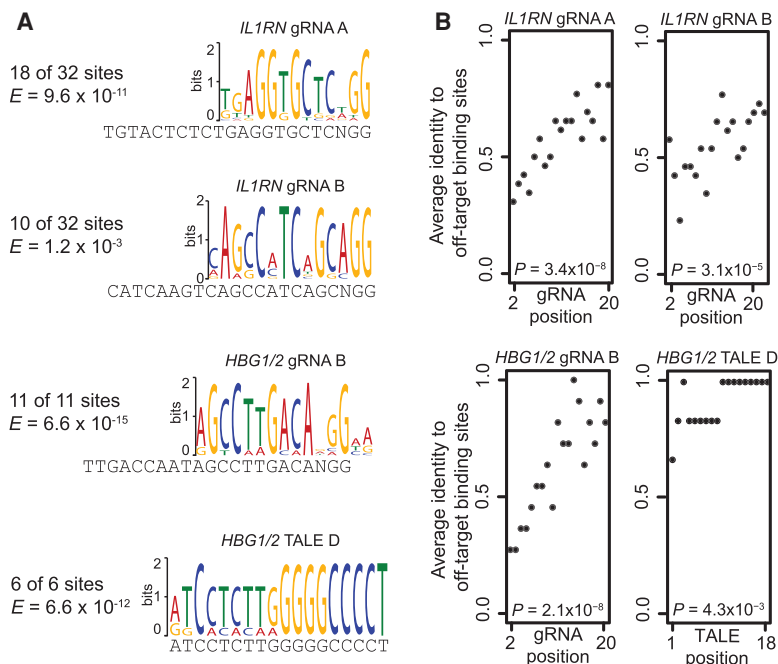
Figure 4. Characterization of dCas9-VP64 and TALE-VP64 off-target binding sites. (*A*) De novo motif detection was used to identify gRNAs and TALEs responsible for identified off-target binding sites. For dCas9-VP64 targeted with gRNAs, motifs matching two of the *IL1RN* gRNAs and one of the *HBG1/2* gRNAs were identified in the respective binding sites identified with ChIP-seq. For TALE-VP64, no motifs matching the *IL1RN* TALEs were identified, and one motif matching a *HBG1/2* TALE-VP64 was identified. (*B*) For each off-target binding site identified by ChIP-seq, we performed an unbiased search for sequences that resemble the intended target sequences of each of the gRNAs or TALE identified in *A*. For this analysis, we considered every possible binding sequence in the called ChIP-seq peak. For dCas9-VP64, we required each possible binding sequence to be followed by the "NGG" PAM sequence. For TALE-VP64, every position in the called binding site was used. Next, for each of the three gRNAs or TALEs identified in *A*, we aligned the intended target sequence to that of every possible binding sequence, and the sequence with the most matching nucleotides in each binding site was retained. DNA sequence similarity to the target sequence at the matched sites was then plotted as a function of the position in the target sequences. For the three gRNAs investigated, a statistically significant trend toward more similarity at the 3′ end of the gRNA sequence was identified, indicating that the 3′ end of the gRNA is more influential in guiding dCas9-VP64 binding. In contrast, the weak 3′ trend observed for TALE D binding is likely an artifact of low sequence complexity in the 3′ end of the target sequence.

we identified significantly enriched motifs with similarity to the sequences targeted by the engineered transcription factors (Supplemental Tables 8–11). In each case, we identified one or two motifs with similarity to the expected target sites (Fig. 4A), but did not find all four motifs. The enrichment of motifs with similarity to particular TALE and gRNA targets, but no enrichment for others, suggests that only a subset of TALEs or gRNAs are responsible for the off-target binding sites. To determine if there is a bias to specific regions of the target sequences in off-target binding events, we aligned the target sequences to all off-target binding sites and calculated the percent identity at every position in the best alignment (Fig. 4B; Supplemental Figs. 1, 2). For gRNA sequences, we restricted the search to locations with canonical PAM sequences. For the gRNA target sequences that we identified with de novo motif identification (*IL1RN* gRNAs A and B and *HBG1/2* gRNA B) (Fig. 4A), we also found a significant increase in similarity to the 3′ end of the same gRNAs ($P < 0.05$, Cochran-Armitage test for trend). Conversely, no trend was observed for the other gRNAs (Fig. 4B; Supplemental Fig. 1). The bias in off-target recognition to the 3′ end of the gRNA is consistent with previous reports that the 3′ end of the gRNA is most important for targeting Cas9 (Cradick et al. 2013; Hsu et al. 2013; Mali et al.

2013a; Pattanayak et al. 2013; Cho et al. 2014; Fu et al. 2014; Sternberg et al. 2014). Therefore, the vast majority of off-target binding sites could be attributed to these three gRNAs, whereas the other five likely had much greater specificity. Meanwhile, we observed no strong or consistent trend in alignments of the sequences that were identified with de novo motif detection to the TALE target sequences in the *IL1RN* promoter (Supplemental Fig. 2), suggesting that the specificity of these TALEs is distributed equally along the length of the array. For TALE-VP64s targeting *HBG1/2*, the de novo motif detection suggested that all the detected off-target binding was the result of a single TALE (TALE D) recognizing sites with an identical GC-rich motif at the 3′ end of the target sequence (Fig. 4).

To determine if off-target binding was leading to changes in gene expression that were not detectable by the genome-wide RNA-seq analysis (Fig. 2), we determined the nearest transcriptional start site to each of the off-target ChIP-seq peaks (Supplemental Tables 12–15). We then compared per-gene mean expression values for these genes, determined by RNA-seq, between control and treated conditions by ANOVA (Supplemental Fig. 3). No significant trend with respect to treatment was observed. Furthermore, we measured the expression of eight of these genes in a targeted manner by the more sensitive qRT-PCR method and did not observe any increases in expression following treatment with the transcriptional activators

(Supplemental Fig. 4). These data, considered together with the low number of ChIP-seq off-target sites (Fig. 3; Supplemental Tables 8–11), the need for multiple TALEs or gRNAs to bind in the same region to effectively alter gene expression (Fig. 1), and our previous observation that genes neighboring *IL1RN* were not affected by *IL1RN*-targeted TALE-VP64s (Perez-Pinera et al. 2013b), collectively provide strong evidence of highly specific gene regulation with undetectable changes in gene expression caused by off-target DNA binding.

We have previously shown that TALE-VP64 and dCas9-VP64 can effectively activate endogenous gene promoters located in closed chromatin (Perez-Pinera et al. 2013a,b), in contrast to previous strategies that focused on targeting open chromatin (Liu et al. 2001; Maeder et al. 2013b,c). To determine whether TALE-VP64 and dCas9-VP64 remodel chromatin structure when activating silenced genes in heterochromatin, we performed DNase-seq to assess genome-wide DNase I hypersensitivity before and after treatment with TALE-VP64 and dCas9-VP64. Compared to cells treated with control plasmid, we detected a substantial increase in chromatin accessibility at the *HBG1* and *HBG2* promoters following transfection with combinations of plasmids encoding either the TALE-VP64s or dCas9-VP64 with gRNAs targeting these
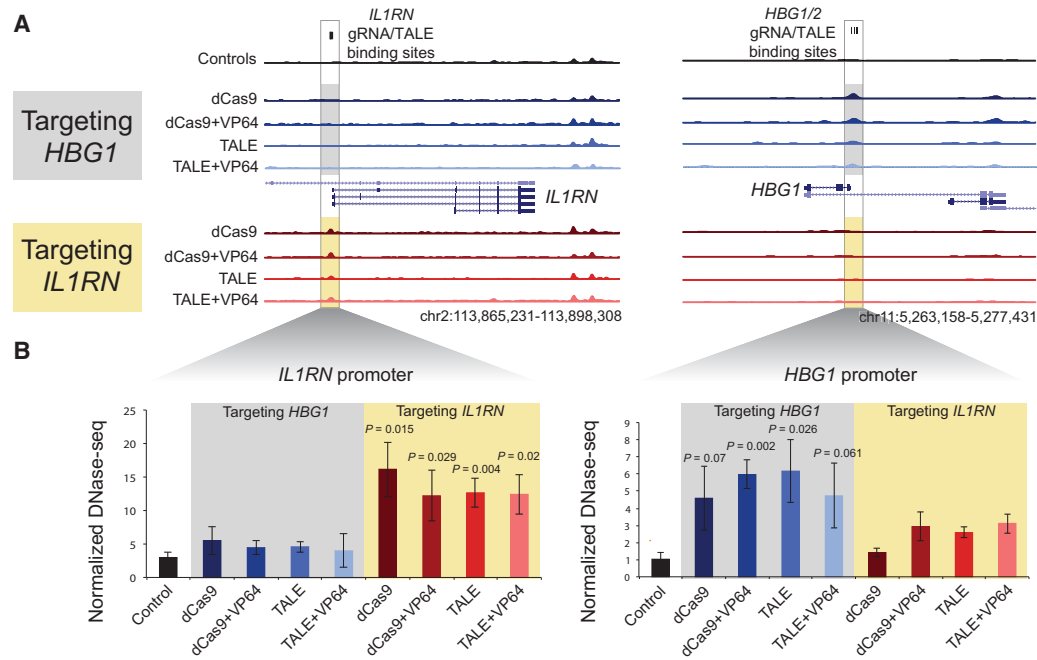
Figure 5. Chromatin accessibility changes induced by both TALEs and dCas9. HEK293T cells were transfected with expression plasmids for TALEs ± VP64 and gRNAs with dCas9 ± VP64 targeted to either the *HBG1/2* promoter or the *IL1RN* promoter. (*A*) Representative DNase-seq data surrounding each promoter (highlighted in box) show increased chromatin accessibility at the promoter to which the TALEs and dCas9 are targeted, but not at the other promoter. (*B*) Normalized DNase-seq cut counts within a 300-bp window surrounding each promoter are shown (mean ± SEM, *n* = 4–6) (Supplemental Table 32). *P*-values are shown compared to the control sample (Tukey's test).

sites (Fig. 5A). We detected a similar increase in chromatin accessibility at the *IL1RN* promoter when using either the TALE-VP64s or dCas9-VP64 with gRNAs targeting this site. When focusing on the 300 bp surrounding the *HBG1* promoter, we observed an increase in normalized DNase-seq cut counts for both TALE-VP64 and dCas9-VP64 targeting the *HBG1* promoter (Fig. 5B). We find the same trend for the TALE-VP64 and dCas9-VP64 targeting the *IL1RN* promoter. Surprisingly, we see the same chromatin remodeling when using either TALEs or dCas9 without the VP64 activation domains, indicating that the changes in chromatin accessibility are not dependent on VP64 and are instead induced by the TALE and dCas9 binding to their target sites (Fig. 5). We find a similar degree of chromatin accessibility changes with TALE or dCas9, indicating that one method is not superior at reconfiguring nucleosome positioning.

To examine the genome-wide specificity of the chromatin remodeling by these proteins, we compared the DNase I hypersensitive sites in samples treated with *IL1RN*-targeted TALEs, with and without VP64, to samples treated with *HBG1/2*-targeted TALEs with and without VP64 (Fig. 6A). We also performed a similar comparison of dCas9 ± VP64 and gRNAs targeted to both genes (Fig. 6B). Although the overall magnitude in change in DNase-seq signal was relatively low (Fig. 5), consistent with the observed moderate levels of RNA-seq signal (Fig. 3; Perez-Pinera et al. 2013a), the results collectively show specific changes in chromatin accessibility at the *IL1RN* and *HBG1*/2 promoters in the expected directions. Assessment of each of the eight treatment conditions individually compared to control similarly showed that the change in DNase hypersensitivity at the target site was one of the most significant across the genome (Supplemental Figs. 5, 6), including the top site for 2/8 conditions, in the top 10 sites in 5/8 conditions, and in the top 70 sites for 7/8 conditions (Supplemental Tables 16–23).

We next tested if off-target changes in chromatin accessibility were associated with changes in expression of nearby genes. Using the same RefSeq annotations used for our RNA-seq analysis, we determined the nearest transcriptional start site for each of the most significant 100 changes in DNase hypersensitivity following treatment with the TALE-VP64s or dCas9-VP64 targeted to *IL1RN* or *HBG1/2* (Supplemental Tables 24–27). There was no significant overall trend in changes in expression of these genes with TALE-VP64 or dCas9-VP64 treatment compared to control (Supplemental Fig. 7). This analysis indicates that any rare off-target changes to chromatin structure generated by TALE-VP64 or dCas9-VP64 do not have significant effects on gene expression.

We next explored whether the off-target TALE-VP64 and dCas9-VP64 binding sites detected by ChIP-seq displayed any significant changes in chromatin accessibility. For each of the ChIP-seq off-target sites shown in Figure 4, we compared DNase-seq signal (Supplemental Tables 28–31) from the control cells treated with empty expression plasmid to cells treated with TALE-VP64s targeted to *IL1RN* (Fig. 6C) and *HBG1/2* (Fig. 6E), and dCas9-VP64 targeted to *IL1RN* (Fig. 6D) and *HBG1/2* (Fig. 6F). Notably, there were no significant changes to DNase-seq signal at any ChIP-seq off-target binding sites, further corroborating the specificity of these tools for transcriptional regulation and indicating that these off-target ChIP-seq sites are not responsible for any detected changes in chromatin accessibility.

We also performed motif analyses to search for subsequences of the gRNA or TALE target sequences in regions of off-target differential DNase I hypersensitive sites. We did not identify any such motifs using the same approach of de novo motif detection that was successfully used to identify enriched motifs in the ChIP-seq data set. We also did not observe any evidence for enriched similarity to the 3′ end of gRNA target sequences. Collectively these
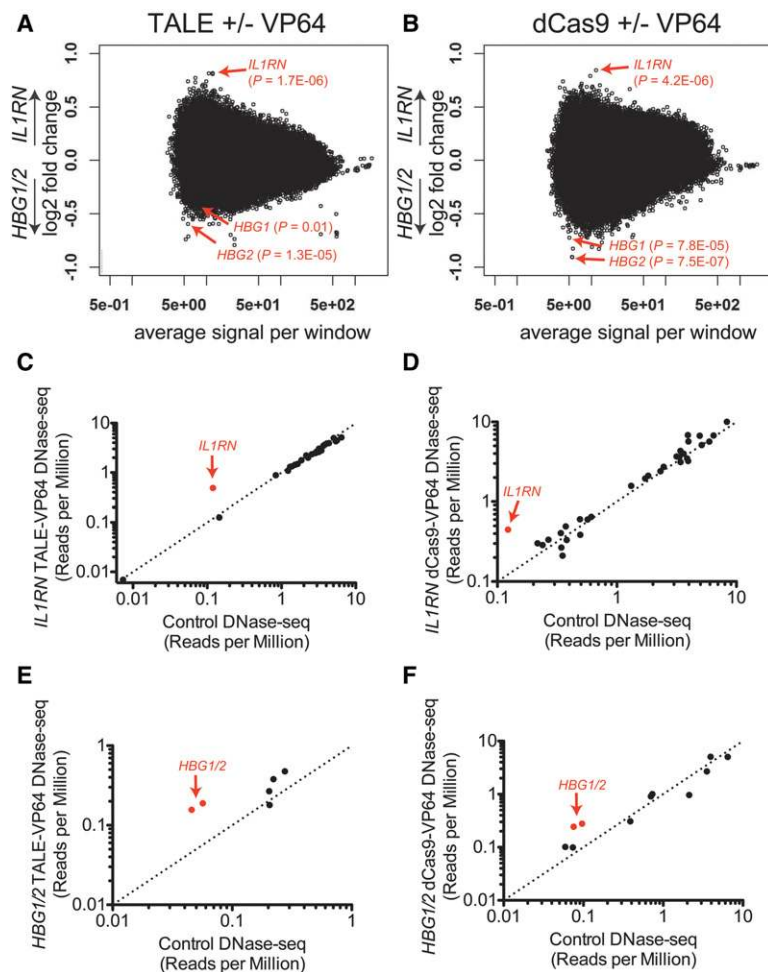
Figure 6. Global characterization of changes to chromatin accessibility. (*A*) Scatter plot of DNase-seq data comparing samples treated with TALEs, with and without VP64, targeted to *IL1RN* versus *HBG1/2*. Each dot represents a DNase I hypersensitive site analyzed by DESeq. *IL1RN* and *HBG1/2* display the expected opposite differences in chromatin accessibility. Nominal *P*-values for each target site are indicated. (*B*) Similar scatter plot as *A*, but for DNase-seq data from *IL1RN*-targeted dCas9 ± VP64 versus *HBG1/2*-targeted dCas9 ± VP64. The individual comparisons of all eight treatments compared to control are presented in Supplemental Figures 5, 6; and the top 100 differential DHS sites for each treatment are provided in Supplemental Tables 16–23. (*C–F*) DNase-seq signal for target (red circles) and off-target ChIP-seq sites (black circles). For each off-target ChIP-seq site, normalized DNase-seq signal from *IL1RN*-targeted TALE-VP64 (*C*), *IL1RN*-targeted dCas9-VP64 (*D*), *HBG1/2*-targeted TALE-VP64s (*E*), and *HBG1/2*-targeted dCas9-VP64 (*F*) was compared to normalized DNase-seq signal from control cells transfected with empty plasmid.

results suggest that the off-target changes to chromatin structure (Fig. 6A,B) were unrelated to TALE or dCas9 activity, and these genome engineering tools are highly specific.

Discussion

Understanding the specificity of genome engineering tools is critical to interpreting results from experiments that intend to measure the effect of only one particular genomic alteration and also to designing therapies targeted to specific genes without causing unwanted side effects. Previous studies of the specificity of TALE- and CRISPR/Cas9-based genome engineering technologies have largely relied on predictive methods, using bioinformatic algorithms (Hockemeyer et al. 2011; Fu et al. 2013; Hsu et al. 2013; Cho et al. 2014; Fine et al. 2014), assays of purified protein activity

in vitro (Pattanayak et al. 2013; Guilinger et al. 2014), and/or reporter assays (Mali et al. 2013a) to inform the selection of off-target sites for direct interrogation. Recently developed methods for unbiased genome-wide specificity analysis are dependent on nuclease activity and therefore cannot directly assess the specificity of gene regulation tools (Frock et al. 2015; Kim et al. 2015; Tsai et al. 2015; Wang et al. 2015). However, other recent studies have used ChIP-seq to directly characterize off-target binding of TALEs or dCas9/gRNAs without nuclease function in mammalian genomes (Mendenhall et al. 2013; Duan et al. 2014; Kuscu et al. 2014; Wu et al. 2014; O'Geen et al. 2015). ChIP-seq has the benefit of agnostic identification of target sites within the cells' genomes. Interestingly, the results between each of these studies and ours vary. For the TALE study, only one target site was identified with no off-target sites (Mendenhall et al. 2013). For the dCas9 studies, hundreds or thousands of off-target binding sites were found for some of the gRNAs, whereas others had as few as 26 off-target sites (Kuscu et al. 2014; Wu et al. 2014). In contrast, we report between four and 31 off-target sites for both TALE-VP64 and dCas9-VP64 (Fig. 3). Importantly, our experiments were performed with pools of four TALE-VP64s and gRNAs, and we showed that some of these individual molecules contributed more to the total number of off-target binding sites than others (Fig. 4). There are many differences in the experimental design of these three studies, including cell type, species, expression system, epitope tag used for ChIP-seq, and whether an effector domain (e.g., VP64) was used. ChIP-seq data analysis also varied, including sequencing depth, filtering, peak calling algorithms, and the requirement for reproducibility across biological

replicates. Regardless, the conclusion from the collective results of all these studies is that both highly specific and highly promiscuous TALEs and gRNAs can be readily identified.

By comparing across eight TALEs and eight gRNAs, our results suggest that these two technologies have similar ranges of genome-wide specificity, which may contribute to alleviating early concerns that CRISPR/Cas9 may be significantly less specific than TALEs or other genome engineering technologies. Importantly, the CRISPR/Cas9 system is generally easier to use compared to other technologies, and it is also easier to find off-target sites for the 20-bp gRNA target site compared to the TALEN dimer that requires two DNA-binding events flanking a spacer of variable length for a total of 30–45 bp of targeted sequence. This may explain why the off-target activity of CRISPR/Cas9 originally gained considerably more attention despite studies clearly showing the potential for off-

target TALE binding and TALEN activity (Hockemeyer et al. 2011; Osborn et al. 2013; Fine et al. 2014; Guilinger et al. 2014). Notably, both technologies showed similar exceptional levels of specificity of gene activation by RNA-seq (Fig. 2; Perez-Pinera et al. 2013a), suggesting that for some applications, these off-target events may be inconsequential, similar to the observation that off-target binding by the Cas9 nuclease frequently does not typically lead to detectable gene editing (Mendenhall et al. 2013; Duan et al. 2014; Kuscu et al. 2014; Wu et al. 2014; O'Geen et al. 2015). The high level of sequence identity in the off-target sites to the on-target sequence (Fig. 4), consistent with other ChIP-seq data for dCas9 (Mendenhall et al. 2013; Duan et al. 2014; Kuscu et al. 2014; Wu et al. 2014; O'Geen et al. 2015), provides support that these are indeed real interactions. However, many of these sites may represent low affinity, short-lived interactions that occur as these proteins search the genome for their perfect target sequence (Sternberg et al. 2014), which is consistent with our observation that the strongest ChIP-seq signal typically is at the intended target site (Fig. 3). If future studies were to identify functional off-target effects of the TALE- and dCas9-based transcriptional regulators, they may be lessened by decreasing the concentration of these proteins inside cells, as has been done for the corresponding nucleases to decrease off-target DNA binding and gene editing (Fu et al. 2013; Hsu et al. 2013; Wu et al. 2014). Additionally, methods for inducible control of these activators have also been developed (Mercer et al. 2014; Polstein and Gersbach 2015; Zetsche et al. 2015).

Interestingly, the ChIP-seq signal was substantially greater for dCas9-VP64 compared to TALE-VP64 for both targets (Fig. 3), despite gene activation by TALE-VP64 being much greater. One potential explanation is that the dissociation of genomic DNA caused by gRNA hybridization leads to disruption of the local DNA conformation and inhibits the action of endogenous transcription factors and regulatory machinery. Although the observation that TALE-VP64 activates genes to a greater extent than dCas9-VP64 has been consistent across several studies and laboratories (Maeder et al. 2013b; Perez-Pinera et al. 2013a; Gao et al. 2014), next generation dCas9-based activator platforms are under development with more robust activity (Chakraborty et al. 2014; Gao et al. 2014; Gilbert et al. 2014; Kabadi et al. 2014; Tanenbaum et al. 2014; Chavez et al. 2015; Hilton et al. 2015; Konermann et al. 2015).

A unique aspect of our study is the observation that both TALEs and dCas9 can be targeted to promoters located in heterochromatin, and the chromatin structure at these target sites is relaxed in response to TALE and dCas9 binding (Fig. 5). The changes to DNase I hypersensitivity were modest, consistent with moderate overall levels of expression of the target gene (Fig. 3). It has previously been shown that TALE-VP64- and dCas9-VP64-mediated gene activation leads to targeted changes in histone modification (Gao et al. 2013, 2014), but the development of strategies for more robust changes to chromatin structure, including the targeted recruitment of histone modifying enzymes (Konermann et al. 2013; Mendenhall et al. 2013; Hilton et al. 2015; Kearns et al. 2015), is an important area of future investigation. This is also supported by observations that TALE-mediated gene activation is facilitated by treatment with inhibitors of DNA methyltransferases or histone deacetylases in some cases (Bultmann et al. 2012).

Typical strategies for activating genes with engineered transcription factors have focused on targeting DNase I hypersensitive sites (Liu et al. 2001; Maeder et al. 2013b,c). This is also consistent with the observation that off-target dCas9 binding is enriched in

open chromatin (Mendenhall et al. 2013; Duan et al. 2014; Kuscu et al. 2014; Wu et al. 2014; O'Geen et al. 2015). However, our results show that targeting strategies do not need to be limited by this restriction, and TALEs and dCas9 may act as pioneer transcription factors with the capacity to activate tightly repressed genes. Surprisingly, chromatin remodeling by these proteins also occurred in response to TALEs and dCas9 without VP64 (Fig. 5), although they were not able to induce activation of the target genes (Fig. 1C). This represents a potential approach to decouple chromatin state and transactivation for fundamental studies of gene regulation.

New methods for improving the specificity of DNA binding by TALE- and CRISPR/Cas9-based tools are rapidly developing (Cho et al. 2014; Fu et al. 2014; Guilinger et al. 2014), and systems with new properties are being engineered (Esvelt et al. 2013). This study provides an outline for unbiased determination of the genome-wide effects of these technologies in the context of transcriptional regulation that will be critical to their advancement in research, medicine, and biotechnology.

Methods

Cell culture and plasmid transfection

HEK293T cells were obtained from the American Tissue Collection Center (ATCC) through the Duke University Cancer Center Facilities and were maintained in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C with 5% CO_2. HEK293T cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transfection efficiencies were routinely >95% as determined by fluorescence microscopy following delivery of a control eGFP expression plasmid. All samples for all assays were harvested at 3 d post-transfection. The dCas9-VP64 expression plasmid was transfected at a mass ratio of 3:1 to either the individual gRNA expression plasmids or the identical amount of gRNA expression plasmid consisting of a mixture of equal amounts of combinations of gRNAs. The expression plasmids for TALE-VP64 (Perez-Pinera et al. 2013b) and dCas9-VP64 and gRNAs (Perez-Pinera et al. 2013a) have been previously described, with the exception that TALE-VP64 expression cassettes targeted to *HBG1/2* were driven by the human ubiquitin C promoter. TALE and gRNA target sequences are provided in Supplemental Table 1. TALEs targeted to the *HBG1/2* promoter were designed using TALE-NT 2.0 (Doyle et al. 2012) and assembled using the GoldenGate kit (Cermak et al. 2011) acquired through Addgene as described previously (Perez-Pinera et al. 2013b).

qRT-PCR

Total RNA was isolated using the RNeasy Plus RNA isolation kit (Qiagen). cDNA synthesis was performed using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Real-time PCR using PerfeCTa SYBR Green FastMix was performed with the CFX96 Real-Time PCR Detection System (Bio-Rad). Primer specificity was confirmed by agarose gel electrophoresis and melting curve analysis. Reaction efficiencies over the appropriate dynamic range were calculated to ensure linearity of the standard curve. Primer sequences and representative standard curves for *IL1RN* and *HBG1/2* have been published previously (Perez-Pinera et al. 2013a). The results are expressed as fold-increase mRNA expression of the gene of interest normalized to *GAPDH* expression and relative to control cells transfected with an equivalent amount of empty expression

plasmid by the $\Delta\Delta C_T$ method. Reported values are the mean and standard error of the mean from two independent experiments performed with biological duplicates on different days ($n = 4$). Statistical analysis was performed by Tukey's test with alpha equal to 0.05 in JMP 10 Pro using log-transformed data to make the variance independent of the mean.

ChIP-seq

ChIP-seq was performed in biological triplicates, where a biological replicate was defined as an independent plate of HEK293T cells transfected on a different day. For each replicate of each condition, ChIP-seq was performed as described previously (Johnson et al. 2007; Reddy et al. 2012). Briefly, for each assay, 20×10^6 transfected HEK293T cells were fixed for 10 min with 1% formaldehyde at room temperature. After quenching the reaction with excess glycine for 5 min, cells were lysed using a solution of 5 mM PIPES, pH 8.0, 85 mM KCl, 0.5% Nonidet P-40, and a protease inhibitor mixture (Roche). The lysate was centrifuged at $300g$ for 5 min at 4°C to collect the intact nuclei. Nuclei were then lysed in RIPA buffer (1× PBS pH 7.4, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, Roche protease inhibitor cocktail), and the chromatin was sonicated using a Diagenode sonicator. Chromatin was immunoprecipitated using a mouse monoclonal antibody targeting the HA tag in dCas9-VP64 or TALE-VP64 proteins (Covance #MMS-101P). After elution, formaldehyde crosslinks were reversed by overnight heating at 65°C, and DNA fragments were purified using a spin column. DNA was then prepared for Illumina high-throughput sequencing using the NEBNext Ultra kit (New England Biolabs #E7370).

In total, 15 ChIP-seq libraries were used in this study. The libraries consisted of three replicates each of five different transfections: dCas9-VP64 + *IL1RN* gRNAs, dCas9-VP64 + *HBG1/2* gRNAs, TALE-VP64 targeting *IL1RN*, TALE-VP64 targeting *HBG1/2*, and empty plasmid. Libraries were sequenced to between 7.7 million and 43 million total reads, and aligned to the hg19 version of the human genome using Bowtie (Langmead et al. 2009) with the "–best" parameter. After alignment, duplicate reads were removed using the SAMtools "rmdup" function. Binding sites were called in each replicate using MACS version 1.4 and relative to a pooled background library consisting of all three ChIP-seq replicates in the cells transfected with empty expression plasmid. Because few off-target sites were identified, we forced MACS to use a shift size between reads aligning to the positive and negative strand of 65 bp rather than trying to build a model de novo. In practice, we did not identify substantial differences between the two approaches. We then identified binding sites that were reproducible across replicates by requiring a pairwise irreproducible discovery rate (IDR) (Landt et al. 2012) <5%. Sites that were reproducible in any pair of replicates were merged into a single list and filtered to remove binding sites identified by the ENCODE Project as likely false positives (i.e., blacklist regions) (Landt et al. 2012). The remaining regions are all of the points (black and red) in Figure 3A–D.

As an additional filter to ensure high-quality binding site calls, we required a statistically significant increase in ChIP-seq signal with dCas9-VP64 or TALE-VP64 relative to empty plasmid-transfected controls according to DESeq (Anders and Huber 2010). To perform that analysis, we counted the number of reads aligned to each candidate binding site in each ChIP-seq replicate for the relevant treatment and for the control. We then calculated read depth normalization coefficients across all binding sites and

used those coefficients to normalize read depth as described previously (Anders and Huber 2010). We then used DESeq to estimate dispersions locally across the pooled counts and then estimated the probability of a greater-than-observed change in read depth between treatment and control (i.e., *P*-values). Finally, a false discovery rate (FDR) was calculated for each site (Benjamini and Hochberg 1995). We considered binding sites with an FDR < 0.1% as our positive set, which are shown as red points in Figure 3A–D.

RNA-seq

RNA-seq libraries were constructed as previously described (Gertz et al. 2012). At 3 d post-transfection, HEK293T cells were lysed using Qiagen RLT-plus buffer with 1% beta-mercaptoethanol. Total RNA was collected using Qiagen RNeasy Plus mini columns including the on-column DNase digestion. Poly-A+ mRNA was isolated from total RNA using a double selection with oligo-dT Dynabeads (Invitrogen), and cDNA was synthesized using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Second-strand synthesis was performed using *E. coli* DNA polymerase I with random hexamer primers (New England Biolabs). Double-stranded cDNA was then collected using Agencourt AMPure XP beads (Beckman Coulter). The Nextera EZ-TN5 transposase was used to simultaneously fragment and insert sequencing primers into the double-stranded cDNA. After 5 min at 55°C, the transposition reactions were halted using Qiagen QG buffer. The fragmented cDNA was then purified using AMPure XP beads. Indexed Illumina high-throughput sequencing libraries were generated by six cycles of PCR. Libraries were constructed for three biological replicates of each condition, for a total of nine libraries.

Libraries were sequenced using 50-bp single-end reads on a single lane of an Illumina HiSeq 2000 instrument. For each transfection condition, reads from the two replicates with the lowest sequencing depth were pooled into a single file, thus creating six data sets with duplicates of each condition. Each data set contained between 9.5 million and 19.6 million reads. Reads were then aligned to human RefSeq transcripts using Bowtie using the "–best" parameter (Langmead et al. 2009). The statistical significance of differential expression, including correction for multiple hypothesis testing, was calculated using DESeq2 (Love et al. 2014).

DNase-seq

DNase-seq libraries were constructed as previously described (Song and Crawford 2010) with the one exception of adding a 5′ phosphate to linker 1 to increase ligation efficiency. Barcoded DNase-seq libraries were sequenced on an Illumina HiSeq 2000, with four barcodes per lane. Replicate number for each treatment and read depth for each sample are provided in Supplemental Table 32. Reads were aligned to human RefSeq using BWA (Li and Durbin 2010), and DNase peaks were called using MACS version 2 (Zhang et al. 2008). For the samples transfected with TALE-VP64 targeted to *HBG1/2*, DNase peaks aligning to the human ubiquitin C promoter were removed manually. Genome-wide statistical significance of differential chromatin accessibility was calculated using DESeq (Anders and Huber 2010).

To identify differential DHS sites, peaks with FDR < 0.01 called by MACS version 2 were identified from each sample and used to generate a union set of DNase HS sites across all samples being compared. From the union set, any DNase HS sites larger than 300 bases were divided into 300-base windows that overlap by 100 bases (Fig. 6A,B). The one exception to this was when comparing each treatment versus control (Supplemental Figs. 5, 6), in

which whole DHS regions from the union set were used. We ensured that the promoter target regions of the *HBG1*, *HBG2*, and *IL1RN* genes were included in the differential chromatin analysis. Raw DNase-seq cut counts for each DNase HS site from each replicate were analyzed by DESeq (Anders and Huber 2010). Genome-wide significance of differential DNase-seq data between experimental conditions (e.g., *HBG1* TALE-targeted lines versus *IL1RN* TALE-targeted lines) was shown by nominal *P*-value calculated by negative binomial distribution, as previously described (Anders and Huber 2010). To make pairwise comparisons of DNase cut counts at the target locus between samples, normalized DNase-seq cut counts within 300-bp windows surrounding the gRNA or TALE binding sites were compared by Tukey's test with α equal to 0.05 in JMP 10 Pro. Similarly, to directly compare DNase-seq signal for the target and all off-target ChIP-seq sites, we extracted raw DNase-seq cut count data from each relevant library that was normalized by the number of total sequences for each library.

De novo DNA-binding motif detection

To identify DNA motifs enriched in sets of binding sites, we used the online MEME software using default parameters and using a range of window sizes (Bailey and Elkan 1994). We limited each binding site identified with ChIP-seq to the 300 bp flanking the predicted point of maximal signal. We then searched for motifs ranging from 4 to 23 bp in length within those binding sites. The 23-bp upper bound is sufficient to recognize the full 20-bp gRNA sequence followed by the 3-bp PAM. We reported five candidate motifs per search. We manually assigned detected motifs to gRNA or TALE target sequences based on sequence similarity.

Data access

The ChIP-seq, RNA-seq, and DNase-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE57085, GSE68341, and GSE67007.

Competing interest statement

Acknowledgments

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28–36.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* **57**: 289–300.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**: 1509–1512.

Bogdanove AJ, Voytas DF. 2011. TAL effectors: customizable proteins for DNA targeting. *Science* **333**: 1843–1846.

Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsaesser J, Lahaye T, Leonhardt H. 2012. Targeted transcriptional activation of silent *oct4* pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* **40**: 5368–5377.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**: e82.

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.

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

Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivalila 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.

Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**: 230–232.

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**: 132–141.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**: 757–761.

Cong L, Zhou R, Kuo YC, Cunniff M, Zhang F. 2012. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**: 968.

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.

Cradick TJ, Fine EJ, Antico CJ, Bao G. 2013. CRISPR/Cas9 systems targeting β-globin and *CCR5* genes have substantial off-target activity. *Nucleic Acids Res* **41**: 9584–9592.

Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, Vandyk JK, Bogdanove AJ. 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40**(**Web Server issue**): W117–W122.

Duan J, Lu G, Xie Z, Lou M, Luo J, Guo L, Zhang Y. 2014. Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res* **24**: 1009–1012.

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

Farzadfard F, Perli SD, Lu TK. 2013. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth Biol* **2**: 604–613.

Fine EJ, Cradick TJ, Zhao CL, Lin Y, Bao G. 2014. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res* **42**: e42.

Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**: 179–186.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**: 822–826.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**: 279–284.

Gaj T, Gersbach CA, Barbas CF III. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**: 397–405.

Gao X, Yang J, Tsang JC, Ooi J, Wu D, Liu P. 2013. Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. *Stem Cell Reports* **1**: 183–197.

Gao X, Tsang JC, 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.

Gersbach CA, Perez-Pinera P. 2014. Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert Opin Ther Targets* **18:** 835–839.

Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, Kuersten S, Myers RM. 2012. Transposase mediated construction of RNA-seq libraries. *Genome Res* **22:** 134–141.

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

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. 2014. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159:** 647–661.

Guilinger JP, Pattanayak V, Reyon D, Tsai SQ, Sander JD, Joung JK, Liu DR. 2014. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* **11:** 429–435.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. 2015. Epigenome editing by a CRISPR-Cas9-based acetyl-transferase activates genes from promoters and enhancers. *Nat Biotechnol* **33:** 510–517.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, et al. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29:** 731–734.

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 nucleases. *Nat Biotechnol* **31:** 827–832.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157:** 1262–1278.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31:** 227–229.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:** 816–821.

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *eLife* **2:** e00471.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316:** 1497–1502.

Kabadi AM, Gersbach CA. 2014. Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods* **69:** 188–197.

Kabadi AM, Ousterout DG, Hilton IB, Gersbach CA. 2014. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res* **42:** e147.

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.

Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim JI, Kim JS. 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12:** 237–243, 1 p following 243.

Konermann S, Brigham MD, Trevino AE, 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.

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.

Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32:** 677–683.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22:** 1813–1831.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589–595.

Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, et al. 2001. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem* **276:** 11323–11334.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE, et al. 2013a. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31:** 1137–1142.

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013b. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10:** 977–979.

Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, Joung JK. 2013c. Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods* **10:** 243–245.

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

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

Meckler JF, Bhakta MS, Kim MS, Ovadia R, Habrian CH, Zykovich A, Yu A, Lockwood SH, Morbitzer R, Elsaesser J, et al. 2013. Quantitative analysis of TALE–DNA interactions suggests polarity effects. *Nucleic Acids Res* **41:** 4118–4128.

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.

Mercer AC, Gaj T, Sirk SJ, Lamb BM, Barbas CF III. 2014. Regulation of endogenous human gene expression by ligand-inducible TALE transcription factors. *ACS Synth Biol* **3:** 723–730.

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29:** 143–148.

Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* **326:** 1501.

O'Geen H, Henry IM, Bhakta MS, Meckler JF, Segal DJ. 2015. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res* **43:** 3389–3404.

Osborn MJ, Starker CG, McElroy AN, Webber BR, Riddle MJ, Xia L, DeFeo AP, Gabriel R, Schmidt M, von Kalle C, et al. 2013. TALEN-based gene correction for epidermolysis bullosa. *Mol Ther* **21:** 1151–1159.

Ousterout DG, Perez-Pinera P, Thakore PI, Kabadi AM, Brown MT, Qin X, Fedrigo O, Mouly V, Tremblay JP, Gersbach CA. 2013. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Mol Ther* **21:** 1718–1726.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31:** 839–843.

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

Perez-Pinera P, Ousterout DG, Brunger JM, Farin AM, Glass KA, Guilak F, Crawford GE, Hartemink AJ, Gersbach CA. 2013b. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods* **10:** 239–242.

Polstein LR, Gersbach CA. 2015. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat Chem Biol* **11:** 198–200.

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.

Reddy TE, Gertz J, Pauli F, Kucera KS, Varley KE, Newberry KM, Marinov GK, Mortazavi A, Williams BA, Song L, et al. 2012. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res* **22:** 860–869.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32:** 347–355.

Song L, Crawford GE. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot5384.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507:** 62–67.

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.

Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, et al. 2015. GUIDE-seq enables genome-

wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33:** 187–197.

Wang X, Wang Y, Wu X, Wang J, Wang Y, Qiu Z, Chang T, Huang H, Lin RJ, Yee JK. 2015. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol* **33:** 175–178.

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482:** 331–338.

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.

Zetsche B, Volz SE, Zhang F. 2015. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33:** 139–142.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29:** 149–153.

Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators

Lauren R. Polstein, Pablo Perez-Pinera, D. Dewran Kocak, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2015/06/03/gr.179044.114.DC1
References	This article cites 82 articles, 14 of which can be accessed free at: http://genome.cshlp.org/content/25/8/1158.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions