CellPress

Molecular Cell
Review

High-Throughput Approaches to Pinpoint Function within the Noncoding Genome

Antonino Montalbano,[1,2,4] Matthew C. Canver,[3,4] and Neville E. Sanjana[1,2,5,*]
[1]New York Genome Center, New York, NY, USA
[2]Department of Biology, New York University, New York, NY, USA
[3]Harvard Medical School, Boston, MA, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: neville@sanjanalab.org
http://dx.doi.org/10.1016/j.molcel.2017.09.017

The clustered regularly interspaced short palindromic repeats (CRISPR)-Cas nuclease system is a powerful tool for genome editing, and its simple programmability has enabled high-throughput genetic and epigenetic studies. These high-throughput approaches offer investigators a toolkit for functional interrogation of not only protein-coding genes but also noncoding DNA. Historically, noncoding DNA has lacked the detailed characterization that has been applied to protein-coding genes in large part because there has not been a robust set of methodologies for perturbing these regions. Although the majority of high-throughput CRISPR screens have focused on the coding genome to date, an increasing number of CRISPR screens targeting noncoding genomic regions continue to emerge. Here, we review high-throughput CRISPR-based approaches to uncover and understand functional elements within the noncoding genome and discuss practical aspects of noncoding library design and screen analysis.

Recent advances in genome engineering tools have enabled targeted interrogation of regions of the human genome that do not code for proteins. These noncoding regions make up nearly 99% of the genome, yet there is little consensus regarding how many of these regions are truly functional (Figure 1A) (Ohno, 1972). Large-scale biochemical studies of noncoding regions, such as the Encyclopedia of DNA Elements (ENCODE) project and Roadmap Epigenomics, suggest that the majority of noncoding DNA is functional (Figure 1B) (Dunham et al., 2012; Roadmap Epigenomics Consortium et al., 2015). Others have examined regions under selective pressure to estimate that ~8% of the genome is functional, while mutational load analysis has suggested that a maximum of 25% of the genome harbors functional regions (Graur, 2017; Rands et al., 2014).

Genome-wide association studies (GWAS) have underscored the importance of noncoding DNA, because >90% of disease- and trait-associated variants are found outside the coding genome (Figure 1C) (Maurano et al., 2012). Identification of clusters of enhancer elements ("super-enhancers") as important regulators of cell identity and disease has further highlighted the significance of regulatory DNA (Whyte et al., 2013). Given the paucity of methods to effectively interrogate the noncoding genome in the past, the path from disease and trait association to the underlying biology has historically been challenging (Edwards et al., 2013).

Despite the importance of noncoding DNA as highlighted by GWAS and enhancer biology, the study of noncoding regions has lagged behind the coding genome. In coding regions, RNA- and protein-level readouts can assess the efficacy of sequence alteration or other experimental interventions. For noncoding regions, no single direct readout exists. However, the advent of genome editing technologies in conjunction with other high-throughput methods such as RNA sequencing (RNA-seq) has provided investigators with new tools for studying the noncoding genome, including promoters, enhancers, introns, microRNAs, long noncoding RNAs (lncRNAs), insulators/repressors, and intergenic regions.

Notably, clustered regularly interspaced short palindromic repeats (CRISPR) genome editing technology offers many strategies for functional interrogation of noncoding DNA. A major advantage of CRISPR-based systems over previous genome editing technologies such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) is their ability to perform high-throughput forward genetic screens rapidly (Shalem et al., 2014; Wang et al., 2014). In a pooled genetic screen, a single investigator can rapidly test hundreds of thousands of genetic hypotheses—a task that would be much more challenging to perform in an arrayed format where each construct is in a separate well. Over the past few years, CRISPR-enabled forward genetic screens have identified key genes involved in diverse aspects of human health, such as cancer, infectious disease, anemias, immune responses to infection, and inborn errors of metabolism (Sanjana, 2017).

Here, we focus on recent advances in high-throughput CRISPR-based approaches to uncover function within noncoding regions. We first review available CRISPR systems and various effector fusions for transcriptional modulation. We then discuss key aspects of using these tools for forward genetic screens, including basics of library design and workflow. We also present recent applications of pooled CRISPR screens in noncoding regions, including specific considerations for data analysis and interpretation.

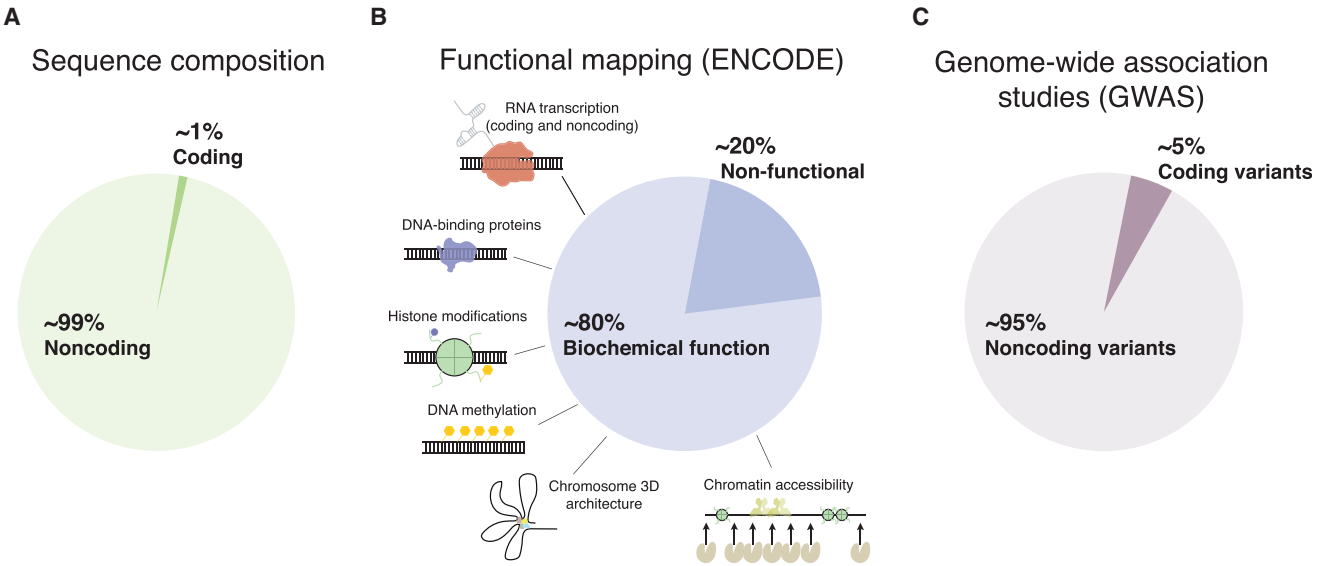

Figure 1. Function in the Noncoding Genome

(A) Distribution of coding and noncoding sequences in the human genome. Noncoding sequences represent the vast majority of the human genome (Dunham et al., 2012).

(B) Estimated biochemically functional portion of the human genome according to ENCODE project mapping (Dunham et al., 2012). This estimate is derived from different biochemical signatures: transcribed regions (protein-coding and noncoding such as enhancer RNAs), regions bound by proteins (e.g., transcription factors), histone modifications associated with function (e.g., histone 3 lysine 27 acetylation, H3K27Ac), DNA methylated regions (methylation of cytosine at CpG dinucleotides), three-dimensional chromosome architecture (gene-distal regions that physically interact and regulate gene expression), and DNase I hypersensitivity sites (regions of open chromatin).

(C) Distribution of GWAS SNPs in the human genome. Most disease-associated SNPs are present in noncoding sequences (Maurano et al., 2012).

CRISPR-Cas Systems for Functional Genomics

CRISPR-Cas systems have been identified as components of the prokaryotic adaptive immune system, and a key hallmark of CRISPR systems is their RNA-guided cleavage of foreign nucleic acids (Barrangou et al., 2007). Specifically, RNA guidance of Cas proteins is used for site-specific cleavage of DNA or RNA. CRISPR-Cas systems are categorized into two classes, which are distinguished by their requirement for multi-protein effector complexes (class 1) or single protein effectors (class 2). Further classification divides each class into types: class 1 includes types I, III, and IV and class 2 includes types II, V, and VI. Notably, there are 19 further classifications into subtypes (e.g., type V-A) (Barrangou and Gersbach, 2017; Koonin et al., 2017). It is likely that classifications of CRISPR-Cas systems will continue to evolve as novel CRISPR-Cas systems are identified.

The diverse CRISPR-Cas systems have a number of features in common, such as the use of short DNA sequences known as "spacers" to direct the targeting of Cas proteins. In addition, there is a requirement for a conserved sequence to aid targeting known as the protospacer adjacent motif (PAM) for DNA-targeted Cas proteins or the protospacer flanking sequence (PFS) for RNA-targeted Cas proteins. Extensive characterization of CRISPR-Cas systems has culminated in its repurposing as a facile platform for eukaryotic genome editing (Cong et al., 2013; Mali et al., 2013a), which has spurred a revolution in biology and medicine (Barrangou and Doudna, 2016).

Class 2 CRISPR-Cas systems have been widely established for eukaryotic genome editing. The initial descriptions utilized the type II effector Cas9 derived from *Streptococcus pyogenes* (SpCas9) (Cong et al., 2013; Mali et al., 2013a). The type II Cas9 requires a CRISPR RNA (crRNA) to be transcribed from the repetitive CRISPR array that contains spacers. The crRNA requires further processing by a *trans*-activating RNA (tracrRNA) before site-specific, blunt-end DNA cleavage can occur upstream of its PAM sequence (NGG) by Cas9's HNH and RuvC catalytic domains. For simplicity, the crRNA and tracrRNA were synthetically fused to create a chimeric single guide RNA (sgRNA) (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013a). With Cas9, the first 20 nucleotides in the sgRNA guide the enzyme to a specific target (which must be adjacent to the PAM sequence) and direct the enzyme to cleave DNA at this location.

Another recently described nuclease validated for genome editing is the Type V-A effector Cas12a (Cpf1) (Zetsche et al., 2015). One distinguishing feature of Cas12a is that it does not require a tracrRNA, unlike type II and V-B systems (Zetsche et al., 2015). Notably, Cas12a allows multiplex targeting by excising multiple crRNAs from a single RNA transcript (Zetsche et al., 2017). Two additional distinguishing features are its cleavage downstream of its TTTV PAM sequence and its creation of 4- to 5-bp 5′ overhangs as opposed to blunt-end cleavage by SpCas9 (Zetsche et al., 2015). The C2c1 nuclease is a type V-B effector, which implies that it requires a tracrRNA in addition to a crRNA. It also cleaves downstream of its TTN PAM sequence and generates 7-bp 5′ overhangs. Interestingly, unlike Cas9 and Cas12a, it does not contain a PAM-interacting domain (Liu et al., 2017a).

Another source of novel nucleases are orthologs of known nucleases from other prokaryotic species. For example, Cas9

orthologs from a variety of different species have been utilized for eukaryotic genome editing. These Cas9 orthologs have diverse PAM sequences: *S. pyogenes* (PAM: NGG), *S. aureus* (PAM: NNGRRT), *S. thermophilus ST1* (PAM: NNAGAA), *S. thermophilus A* (PAM: NGGNG), *N. meningitides* (PAM: NNNNGATT), *C. jejuni* (PAM: NNNNRYAC), *B. laterosporus* (PAM: NNNNCNDD) (Karvelis et al., 2017). Another strategy to expand the CRISPR toolkit has been the use of directed evolution and/or structural information-guided mutagenesis to alter PAM specificity. For example, SpCas9 has been altered through three amino acid changes to the SpCas9-VQR variant with NGA PAM specificity, while four amino acid changes to SpCas9 resulted in the SpCas9-VRER variant with NGCG PAM specificity (Kleinstiver et al., 2015a). Similarly, three amino acid substitutions in the PAM-interacting domain of *S. aureus* (SaCas9) altered its PAM specificity from NNGRRT to NNNRRT (SaCas9 KKH-variant) (Kleinstiver et al., 2015b). Beyond altered PAM Cas9 nucleases, Cas12a has recently been engineered with different PAMs: TYCV with two and TATV with three amino acid substitutions, respectively (Gao et al., 2017). PAM sequences have also been identified for two recently discovered class 2 nucleases, CasX (PAM: TTCN) and CasY (PAM: TA) (Burstein et al., 2017). Although most studies have focused on class 2 effectors, recent work on type I-E CRISPR-Cas (Cascade) and Cas3 nuclease revealed a PAM specificity of NNNAAG (Xiao et al., 2017). Lastly, recent reports have described CRISPR-Cas systems with RNA-guided RNase activity. These include the type VI effectors Cas13a (C2c2) (Abudayyeh et al., 2016; East-Seletsky et al., 2016) and Cas13b (C2c6) (Smargon et al., 2017). RNA-targeted Cas proteins are beyond the scope of this review but hold tremendous promise for targeted manipulation of the transcriptome.

Applications of CRISPR-Cas Technology

In the few years since the initial description of CRISPR technology for eukaryotic genome editing, applications for programmable nucleases have expanded rapidly. This includes methods for knockout, knockin, interference, activation, base editing, and epigenome editing (Figure 2) (Barrangou and Doudna, 2016). These applications fall into two categories: those that alter the targeted DNA sequence, and those that do not alter the targeted DNA sequence. In general, use of nucleases with intact catalytic activity will result in a double-strand break (DSB) (Figure 2A). The DSB is primarily repaired by either non-homologous end joining (NHEJ) repair or homology-directed repair (HDR). NHEJ is the repair process of choice for mutagenesis of a genomic region or gene knockout because NHEJ typically results in a small (1–10 bp) insertion-deletion mutation (indel) (Cong et al., 2013; Mali et al., 2013a). HDR is a templated repair process that can be harnessed to insert a custom sequence into the genome by using an exogenous repair template. HDR efficiency in mammalian cells lags behind NHEJ efficiency; however, a greater understanding of the pathways involved in stimulating HDR and its regulation by the cell cycle holds promise for improvement.

Base editing approaches represent an alternative to a reliance on HDR to edit genomic sequences without creating a DSB. Base editing can be performed by the fusion of a cytidine deaminase enzyme, such as rat-derived APOBEC1, to catalytically inactive Cas9 (dCas9) (Komor et al., 2016) (Figure 2B). Other cytidine deaminase enzymes, such as activation-induced cytidine deaminase (AID), have also been used (termed "CRISPR-X") (Hess et al., 2016; Nishida et al., 2016). These fusions allow for modification of C→T or G→A in the absence of a DSB, with mutation predominantly occurring in a 3- to 5-bp window, which was recently narrowed to 1 to 2 bp for enhanced targeting specificity (Kim et al., 2017c). The mutation profile was recently shown to be expandable by using an AID-P182X (AIDx) mutant, which can convert C and G to the other three nucleotides at appreciable frequencies (Ma et al., 2016). Base editing methods have been shown to be enhanced by using a Cas9 nickase as well as a DNA glycosylase inhibitor protein (e.g., UGI), which has allowed for editing rates of 15%–75% (Komor et al., 2016; Nishida et al., 2016). Genome-wide specificity of base editing approaches continues to be evaluated and improved (Kim et al., 2017a).

CRISPR interference (CRISPRi) is a method used for the repression of transcription akin to RNA inference (RNAi) (Figure 2C). CRISPRi blocks transcription itself instead of degrading the transcript through RNAi. CRISPRi utilizes a catalytically inactive nuclease, such as dCas9. Targeting of a catalytically inactive nuclease results in transcriptional repression by steric hindrance as well as interfering with the binding of transcription factors and/or RNA polymerase binding and processivity (Qi et al., 2013). In mammalian systems, dCas9 itself is not sufficient for repression, and effective repression has required a fusion of dCas9 with the Krüppel-associated box (KRAB) effector domain (Gilbert et al., 2013). KRAB repression is mediated by spreading of repressive histone modifications such as H3K9me3. Beyond KRAB, several other epigenome-modifying repressor fusions with CRISPR proteins have been utilized, including Lys-specific histone demethylase 1 (LSD1) (Kearns et al., 2015), histone deacetylase (HDAC) (Kwon et al., 2017), DNA methyltransferases DNMT3A and MQ1 (Lei et al., 2017; Liu et al., 2016; Vojta et al., 2016), and other repressive domains previously used with ZFNs/TALENs (Konermann et al., 2013; Thakore et al., 2016). To date, no systematic comparison of different CRISPRi systems has been performed in mammalian systems.

For gene activation, catalytically inactive nucleases can be fused to effector domains for transcriptional activation (termed CRISPRa) (Cong et al., 2013; Maeder et al., 2013; Mali et al., 2013b) (Figure 2D). Initial CRISPRa studies involved the fusion of single activation effectors, such as VP64 (Mali et al., 2013b) or VP160 (Cheng et al., 2013), but activation did not work for all genes and often required targeting several regions of the promoter simultaneously. Subsequent studies demonstrated the use of multiple activation effectors together, such as MS2-p65-HSF1 with dCas9-VP64 ("SAM") (Konermann et al., 2015) and dCas9-VP64-p65-Rta ("VPR") (Chavez et al., 2015), and peptide arrays ("SunTag") (Tanenbaum et al., 2014). A recent comparison of these methodologies demonstrated that SAM, VPR, and SunTag were superior to VP64 alone; however, the study was unable to reliably differentiate between the efficiency of the three approaches (Chavez et al., 2016). In addition to fusion of activation domains, studies have adopted further modifications such as the addition of MS2 bacteriophage coat protein-binding aptamers

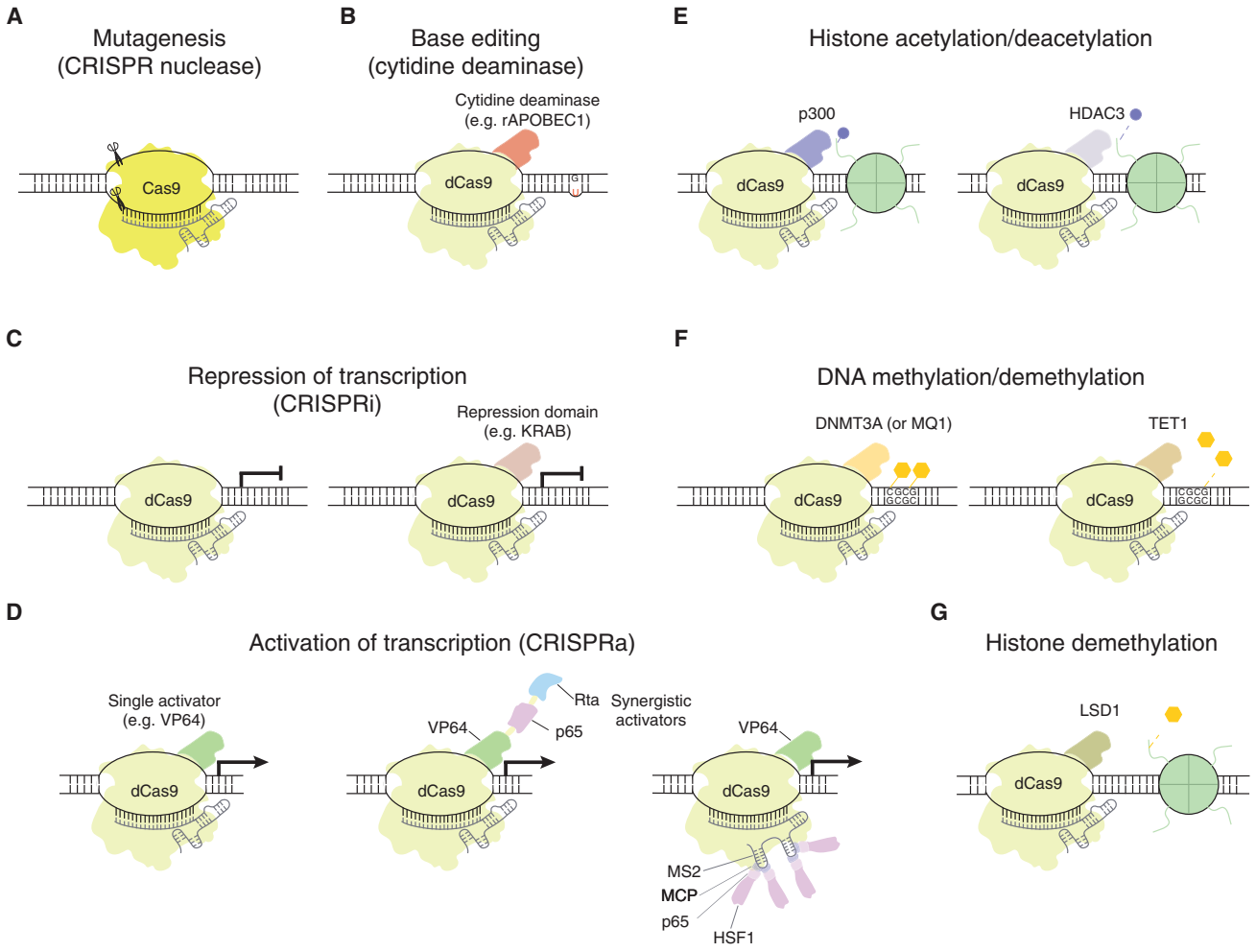

Figure 2. Different Pooled CRISPR Screening Modalities

(A) CRISPR nucleases such as Cas9 can be used to disrupt coding and noncoding regions by making use of NHEJ to introduce insertion or deletions (indels) in the sequence of interest (CRISPRn) (Cong et al., 2013; Mali et al., 2013a).

(B) Catalytically inactive Cas9 (dCas9) can be fused to cytidine deaminase (e.g., rat APOBEC1) to introduce single-nucleotide C-to-T transitions in the target sequence (Komor et al., 2016).

(C) dCas9 can be applied to achieve transcriptional repression (CRISPRi). dCas9 can physically repress transcription through steric hindrance (Gilbert et al., 2013; Qi et al., 2013) or, alternatively, can be fused to repressor domains, e.g., the Krüppel-associated box (KRAB) domain (Gilbert et al., 2013).

(D) dCas9 can be used to activate transcription through different strategies (CRISPRa). Examples are dCas9 fused to the Herpex simplex virus protein 64 (VP64) transcriptional activator alone (Mali et al., 2013b) or combined with other activators, e.g., p65 and Epstein-Barr virus R transactivator (Rsa) (Chavez et al., 2015), modified gRNA-MS2 that recruits MCP fused to additional activators such as p65, and heat shock factor 1 (HSF1) (Konermann et al., 2015).

(E–G) The CRISPR system can also be used to introduce epigenetic modifications.

(E) dCas9 fused to p300 (Hilton et al., 2015; Klann et al., 2017) or histone deacetylase 3 (HDAC3) (Kwon et al., 2017) can perform targeted acetylation or de-acetylation to activate or inhibit transcription, respectively. Purple circles indicate the acetyl group.

(F) dCas9 fused to DNA methylases (e.g., DNMT3A) (Lei et al., 2017; Liu et al., 2016; Vojta et al., 2016) or demethylases (TET1) (Liu et al., 2016) introduces or removes methyl groups at CpG dinucleotides (silencing of gene expression), respectively. Orange hexagons in (F) and (G) indicate methyl groups.

(G) dCas9 fused to lysine-specific histone demethylase 1 (LSD1) (Kearns et al., 2015) catalyzes the demethylation of histone 3 lysine 4 and 9 (H3K4 and H3K9), resulting in transcriptional repression.

to guide RNA (gRNA) sequences aimed to enhance recruitment of activation domains. These MS2 aptamers have been added to the 3′ end, tetraloop, and stem loop 2 (Konermann et al., 2015; Mali et al., 2013b). Epigenome editing approaches can also be used for targeted transcriptional activation, such as dCas9 fusion to a DNA demethylase (e.g., TET1) or to a histone acetyltransferase (e.g., p300) (Hilton et al., 2015; Klann et al., 2017; Liu et al., 2016). Conversely, transcriptional repression can be mediated by dCas9 fusion to either a DNA methyltransferase (e.g., DNMT3A, MQ1) or a histone demethylase (e.g., LSD1) (Figures 2E–2G) (Kearns et al., 2015; Lei et al., 2017; Liu et al., 2016; Vojta et al., 2016).

Pooled Screening Strategies

Given the growing CRISPR toolbox, a variety of high-throughput pooled screening options have become available for

nuclease-driven mutagenesis, CRISPRi, CRISPRa, epigenome modification, and base editing (Figure 2). Pooled screens typically combine these genetic manipulations with positive, negative, or marker/reporter gene selection. Common examples of selection include fluorescence-activated cell sorting (FACS), drug or toxin sensitivity/resistance, and cell survival and proliferation. After selection, pooled screens are analyzed by counting gRNAs to calculate the enrichment or dropout ("depletion") between conditions (e.g., drug treatment versus vehicle, early time point versus later time point, reporter gene positive versus reporter gene negative). The goal is to identify which gRNAs are enriched (or depleted) in the comparison.

In addition to traditional selection paradigms, single-cell RNA-seq has recently been developed for pooled CRISPR screen readout as an alternative strategy ("Perturb-seq"), allowing capture of multi-dimension transcriptional phenotypes to correlate with the genetic manipulation (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016). These studies will likely pave the way for pairing of pooled screening with a variety of other single-cell techniques, such as single-cell Hi-C (Ramani et al., 2017), whole-genome sequencing (Vitak et al., 2017), DNA methylation (Guo et al., 2013; Smallwood et al., 2014), ChIP-seq (Rotem et al., 2015), proteomics (Stoeckius et al., 2017), or Assay for Transposable and Accessible Chromatin (ATAC) sequencing (Buenrostro et al., 2015; Cusanovich et al., 2015). Combinations of these readouts are also possible: one recent study reported simultaneous single-cell measurement of chromatin accessibility, DNA methylation, and nucleosome phasing (Pott, 2017).

Pooled knockin screens using HDR are technically feasible but have not been widely adopted. This is likely due to low observed rates of HDR as compared to NHEJ mutagenesis (Cong et al., 2013; Mali et al., 2013a). One example study performed a pooled HDR screen to knock in all possible variants into a 6-bp region of exon 18 of the *BRCA1* gene (Findlay et al., 2014). Another study exploited HDR in a pooled format for knockin of library gRNAs as an alternative to standard viral delivery (Rajagopal et al., 2016). As HDR rates continue to improve through further understanding of factors affecting HDR, pooled knockin screens may be more fully realized.

For loss-of-function screens, both CRISPR mutagenesis and CRISPRi approaches have been used. One study directly compared a CRISPR knockout/mutagenesis screening approach to a CRISPRi screen targeting essential genes and cancer-associated genes (Rosenbluh et al., 2017). This resulted in CRISPR knockout identifying 98% of essential genes as compared to 92% for CRISPRi. The authors further noted that using CRISPR mutagenesis typically resulted in "stronger" phenotypes and a greater proportion of effective gRNAs as compared to using CRISPRi; however, they also noted that both techniques suffer from risks of false positives and negatives (Rosenbluh et al., 2017). For CRISPR mutagenesis, false positives can result from amplified genomic regions, where multiple DSBs induce apoptosis (Aguirre et al., 2016; Munoz et al., 2016), and false negatives can result from exon skipping (Mou et al., 2017). Given that the majority of mutagenesis screens have been performed in cancer cell lines, it is important to keep this consideration in mind when targeting amplified regions such as those near oncogenes. In contrast, false positives can result from bidirectional promoters, and false negatives can result from gene expression driven by multiple promoters for CRISPRi (Rosenbluh et al., 2017).

Pooled screens have also enabled high-throughput genetic studies *in vivo*, which can include viral infection of CRISPR reagents *ex vivo* with subsequent transplantation (Chen et al., 2015; Manguso et al., 2017) or direct delivery of CRISPR reagents *in vivo* (Chow et al., 2017). Delivery of both the gRNA and nuclease in a single vector is preferable for *in vivo* applications to achieve higher transduction rates as opposed to a split vector system (i.e., separate vector for nuclease and gRNA) requiring co-transduction (Ran et al., 2015; Xue et al., 2014). Others have avoided this technical issue by using constitutive or inducible Cas9-expressing mouse models, which only require gRNA delivery (Platt et al., 2014; Wang et al., 2017). The use of a Cas9-expressing mouse can facilitate pooled *in vivo* screening due to the higher titer of vectors containing only gRNAs in contrast to vectors with both the gRNA and nuclease (Sanjana et al., 2014; Wang et al., 2017).

Library Design

Library design requires identification of loci for targeting, choice of CRISPR nuclease or effector fusion, and appropriate positive and negative controls. Logically, the type of nuclease or effector also influences the library design strategy. For example, CRISPR knockout screens typically target gene exons, whereas transcriptional modulation screens using CRISPRa or CRISPRi usually target gene promoters. Also, regardless of the effector domain, targeting requirements for noncoding regions can be different from those for coding regions.

Previous work to knock out protein-coding genes has focused on targeting exons that are constitutively expressed, avoiding the last exon to prevent escape from nonsense-mediated transcript decay, and targeting protein functional domains. The increased knockout with functional domain targeting is likely due to the fact the even in-frame indel mutations are deleterious in functional domains (Shi et al., 2015). Saturating mutagenesis screens over protein-coding genes have also elucidated several base preferences for the Cas9 nuclease to enable the design of gRNAs with higher on-target cleavage activity (Doench et al., 2014, 2016). Software tools are available that integrate several of these criteria for the design of custom CRISPR libraries (Meier et al., 2017).

Notably, CRISPRi and CRISPRa pooled screens have specific requirements for targeting near the transcription start site (TSS). A screen was performed tiling gRNAs within a 10-kb window around 49 genes (n = 54,810 gRNAs) to identify the optimal targeting region for effective CRISPRi. CRISPRi-mediated repression peaked in close proximity to the TSS, with the strongest effect between −50 and +300 bp of the TSS (Gilbert et al., 2014). CRISPRa also relies on targeting in close proximity to the TSS, with the SAM activation system being most effective between −200 and 0 bp from the TSS (Konermann et al., 2015). Since many genes have alternative 5′ splice isoforms, TSS determination can also be a challenge: for CRISPRi repression, the FANTOM/CAGE promoter atlas has been shown to provide reliable TSS annotations (Radzisheuskaya et al., 2016). When targeting lncRNAs, mutagenesis within the lncRNA body may not result in functional ablation. To determine an optimal

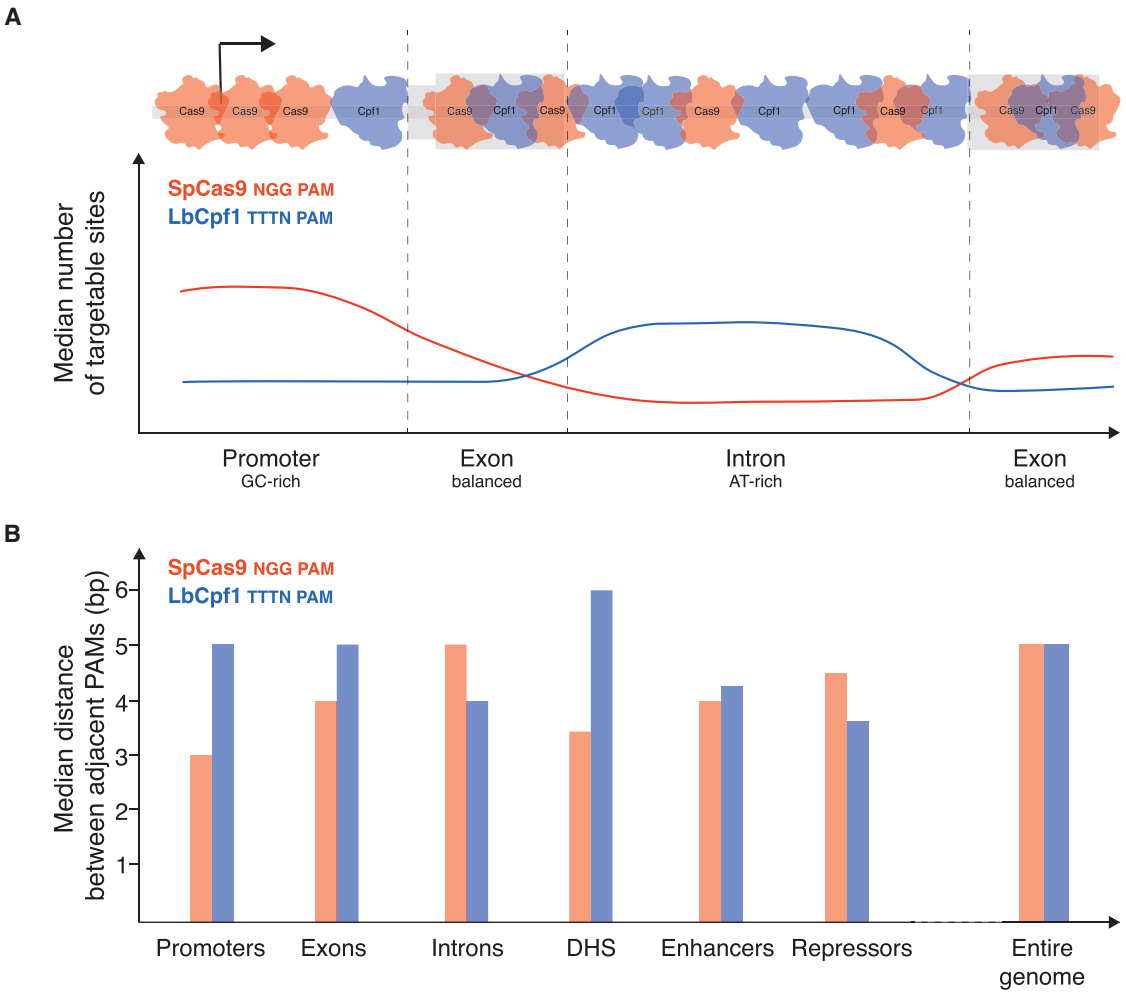

Figure 3. PAM Distribution across the Human Genome

(A) Schematic of the distribution of GC-rich and AT-rich protospacer-adjacent motifs (PAMs) in a single gene. Distribution of target sites/PAMs is computed as the inverse of the median distance between PAMs for SpCas9 (NGG) and LbCpf1 (TTTN). The PAM distribution of a particular nuclease is an important parameter to consider when targeting specific noncoding elements.

(B) Median distance (in bp) between PAMs of SpCas9 (NGG) and LbCpf1 (TTTN). Although SpCas9 and LbCpf1 PAMs display an even distribution across the human genome, the SpCas9 PAM is more frequent in GC-rich regions such as promoters, whereas the LbCpf1 PAM (TTTN) is more represented in AT-rich regions such as introns (Canver et al., 2017a).

CRISPRi strategy for lncRNA repression, one study tiled gRNAs across the gene body of the lncRNA *PVT1* (17,469 gRNAs); it found that targeting within 1 kb of the most upstream TSS of the lncRNA was effective, whereas targeting exons and introns was ineffective (Liu et al., 2017b).

Other considerations for library design for nucleases and effector fusions include utilization of engineered PAM domains for higher tiling density and variant-aware design for outbred organisms (e.g., human) where differences from the reference genome may be substantial (Canver et al., 2017a). Different regions of the genome can have substantial differences in base composition. For example, the T-rich PAM of Cpf1 is more abundant in introns, whereas the G-rich PAM of SpCas9 is more abundant in exons (Figure 3). Evaluation of the median distance between adjacent genomic cleavages can be used to evaluate degree of saturation of PAMs within a given genomic region (Figure 3B). Once a nuclease is chosen, it is important to ensure that gRNA target sequences are designed using the optimal length for that nuclease. For example, 20 bp is typically used for SpCas9 (Hsu et al., 2013), 22 bp for CjCas9 (Kim et al., 2017b), and 21–23 bp for SaCas9 (Ran et al., 2015).

Negative and positive controls should also be included in the design of the screen. Negative controls are essential for library design to evaluate experimental noise and the effects of the delivery of CRISPR reagents. Published libraries have often included non-targeting controls as 1%–5% of the total number of gRNAs in the library (Sanjana et al., 2014), but recent studies have suggested that gRNAs targeting non-essential genes can also serve this purpose (Hart et al., 2017). Alternative negative controls include targeting safe-harbor genomic regions (e.g., *AAVS1*) (Rosenbluh et al., 2017; Wang et al., 2014) as well as targeting GFP or luciferase in GFP- and luciferase-negative cells,

respectively. Positive controls are also essential to ensure validity of screen data. Similar to negative controls, positive controls can be included at the library design stage or separately synthesized and spiked into relevant libraries. However, unlike negative controls, positive controls are rarely generalizable and are typically experiment-specific. For gene-targeted screens, positive controls should target genes previously known to be associated with a given phenotype. Targeting the coding sequence of the gene of interest can serve as a positive control in noncoding screens (Canver et al., 2015, 2017a; Sanjana et al., 2016). Alternatively, targeting GFP can serve as a positive control in pooled screen experiments using a GFP reporter-based readout (Diao et al., 2017; Rajagopal et al., 2016), and targeting universally essential genes, such as ribosomal genes, can serve as a positive control for dropout screens (Shalem et al., 2014; Wang et al., 2014).

Pooled CRISPR Screen Workflow

Most pooled CRISPR screens follow a similar workflow (Figure 4): (1) design and synthesis of the gRNA library, (2) viral transduction into target cells, (3) selection of a phenotype relevant to the disease or biological process to interrogate, (4) readout by next-generation sequencing of the gRNAs after selection, and (5) enrichment or depletion analysis of gRNAs. Although not part of the pooled screen itself, it is vital to validate top candidates identified in the pooled screen using a separate arrayed screen (preferably using gRNAs targeting similar genes and/or elements that were not in the pooled library) and other orthogonal and complementary biological assays. Several detailed step-by-step protocols for pooled CRISPR screens have been published recently (Canver et al., 2017b; Joung et al., 2017a; Wang et al., 2016).

For library design, the first choice is whether to use an existing gRNA library or a custom-designed library. For protein-coding genes, libraries designed for nuclease, CRISPRi, or CRISPRa screens are available from multiple labs through Addgene. Libraries targeting noncoding regions or particular genes are usually designed using custom software to pull out genomic regions of interest and identify gRNAs within those regions. Once the *in silico* design is complete, gRNAs are synthesized as DNA oligonucleotides for pooled cloning. Studies to date have most commonly cloned gRNA libraries into viral vectors such as lentiviral or AAV vectors (Shalem et al., 2014; Wang et al., 2017, 2014). A number of strategies have been described for cloning pooled CRISPR libraries such as Gibson assembly, which uses PCR amplification to make double-stranded DNA with 20- to 30-bp overhangs from synthesized single-stranded oligonucleotides (Shalem et al., 2014; Wang et al., 2014). For deletion screens with SpCas9 and multiple gRNAs, each gRNA is under the control of a different promoter (U6 and H1 promoters or human U6 and murine U6) and thus requires multiple pooled cloning steps. It is essential to sequence the plasmid pool to ensure successful cloning of the library regardless of the chosen cloning method prior to screen experiments.

To construct the plasmid pool, amplified library gRNAs are typically cloned into viral vectors for precise control over the multiplicity of infection, which prevents cells from receiving multiple gRNA constructs. Initial CRISPR screens delivered both nuclease and gRNA in the same vector, which often results in low viral titer given the large size of most CRISPR enzymes (Sanjana et al., 2014). As an alternative, cells with stable nuclease expression need only delivery of a gRNA in the viral vector. Although a heterogeneous population of nuclease-expressing cells can be used, it is preferable to select a stable clone based on nuclease expression (by western blot, qRT-PCR, or marker gene expression) (Chen et al., 2015). This approach of using cells with stable nuclease expression may not be possible in certain cases, such as experiments using primary cells with limited culture duration. After viral transduction, cells are selected via drug or other markers to keep only those cells that received a gRNA. Generally, cells are cultured for 1–2 weeks before beginning phenotype-based selection so that there is sufficient time for gene editing and downstream changes in transcript and protein levels.

To readout the screen, the representation of the library after phenotypic selection is compared with the representation at an early time point. Libraries are prepared for deep sequencing by PCR from genomic DNA using primers specific to the genomically integrated CRISPR construct or using locus-specific primers for non-integrating library delivery methods, such as AAV (Chow et al., 2017; Wang et al., 2017). Sequencing reads are aligned to the original gRNA library, resulting in read counts for each gRNA in the library before and after phenotypic selection. Typical data analysis examines the consistency among different gRNAs targeting the same gene or noncoding element (see Analysis of Pooled CRISPR Screens below). Before discussing analysis techniques, we will review noncoding screens targeting regions such as super-enhancers, enhancers, DNase-hypersensitive sites (DHSs), binding motifs, and noncoding RNAs.

Pooled CRISPR Screens

Protein-Coding Genes

Although gene-targeted pooled screens are not the focus of this review and have been reviewed elsewhere (Shalem et al., 2015), many of the concepts that have been applied to gene-targeted pooled screens are applicable to noncoding screens. The majority of pooled CRISPR screens have used the Cas9 nuclease for gene knockout and have been performed at genome-wide scale (Koike-Yusa et al., 2014; Shalem et al., 2014, 2015; Wang et al., 2014). Other gene-targeting strategies have included paired gRNAs to target two genes in a single cell (Han et al., 2017; Shen et al., 2017; Wong et al., 2016), mutagenesis to identify gain-of-function mutations resulting from in-frame NHEJ repair (Donovan et al., 2017), base editing to introduce stop codons to knock out genes (Kuscu et al., 2017), protein domain targeting (Shi et al., 2015), and *in vivo* screening for phenotypic selection not possible to perform *in vitro* (Chen et al., 2015; Xu et al., 2017). In addition to mutagenesis with Cas9 nuclease, genome-wide screens have also been performed with CRISPRa/i systems for gene activation/repression (Gilbert et al., 2014; Konermann et al., 2015).

Saturating Mutagenesis to Identify Noncoding Functional Elements

Several noncoding CRISPR screens have utilized saturating mutagenesis in putative enhancer regions (identified by biochemical marks like DHS and H3K27Ac) or in larger regions

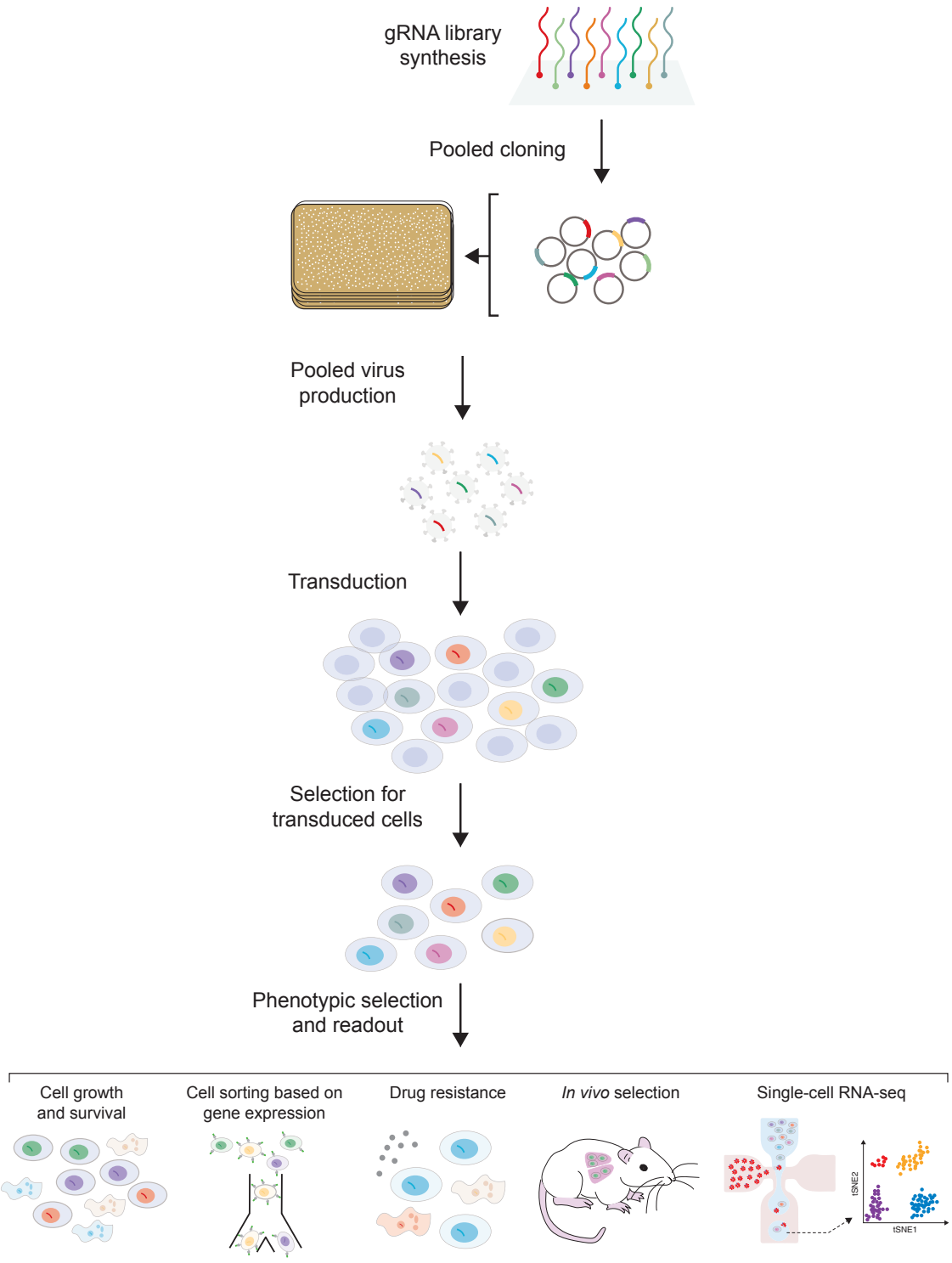

Figure 4. Pooled CRISPR Screen Workflow and Phenotypic Selection

In pooled CRISPR screens, gRNAs are synthesized, cloned, and constructed as a pool. Typically, viral transduction is performed at a low multiplicity of infection so that each cell receives one viral particle. Viral integration into the genome enables amplification of the gRNA cassette through PCR and readout through next-generation sequencing. During readout, the abundance of the different gRNA cassettes is quantified and a differential analysis of gRNA abundance between samples is performed. For example, gRNAs targeting coding or noncoding regulatory regions of genes that are essential for cell survival will drop out in a pooled screen (Shalem et al., 2014; Wang et al., 2014). In screens targeting noncoding regions that regulate transcription of the gene of interest, gRNAs that

(legend continued on next page)

adjacent to phenotypically relevant genes to pinpoint enhancers. Nuclease-driven saturating mutagenesis attempts to saturate a region with indel mutations to identify functional noncoding sequence. The key principle is that indels at the site of the functional noncoding element will abrogate gene regulatory function. Noncoding screens have used a variety of different strategies to design gRNA libraries for targeting enhancer regions.

An early study targeted an ~10-kb enhancer of *BCL11A*, a region implicated by GWAS to modulate fetal hemoglobin levels and thus therapeutically significant for diseases such as sickle-cell anemia and β-thalassemia (Canver et al., 2015). Noncoding gRNA libraries tiled the human enhancer and its mouse homolog with ~500 gRNAs each to identify which regions within the ~10 kb enhancer were responsible for modulating BCL11A and fetal hemoglobin. For each library (human and mouse), positive controls targeting *BCL11A* exons and negative controls that do not target in the respective genomes were included. The degree of saturation was evaluated by determining the distance between adjacent genomic sequences. For example, the median spacing was 4 bp for both the mouse and human libraries. After genome modification, cells were selected via FACS-based selection for fetal hemoglobin expression. A peak of enriched gRNAs led to the identification of a narrow region that includes a GATA1 binding site as essential for *BCL11A* expression.

A different approach involves targeting regions near a gene of interest, including introns, untranslated regions, promoters, and intergenic sequences. This targeting strategy can be performed with or without guidance from GWAS or previous enhancer mapping. To detect gene regulatory elements in the vicinity of a gene, a recent study tiled gRNAs across ~100-kb regions both upstream and downstream of coding exons around three genes involved in kinase inhibitor resistance in melanoma (Sanjana et al., 2016). In total, nearly a megabase of the noncoding genome was targeted. This study identified a predominance of enriched gRNAs upstream of gene bodies as opposed to downstream, and more enriched gRNAs near genes that harbor greater numbers of expression quantitative trait loci (eQTLs). Enriched gRNAs were also found to colocalize with open chromatin as assessed by ATAC-seq and DHS analyses and, through chromosome conformation capture (3C), were more likely to target regions that are in close proximity to the gene promoter. Perhaps most intriguingly, this work identified several functional elements that exist outside of traditional biochemical hallmarks of regulatory function, suggesting that noncoding screens can identify functional elements that may not be found through other assays. Interestingly, enriched gRNAs were located in close proximity to regions conserved in primates (but not those conserved in vertebrates or mammals), a finding similar to the *BCL11A* study results wherein functionally similar mouse and human enhancers were found in distinct genomic locations (Canver et al., 2015). Taken together, these saturating mutagenesis experiments are consistent with enhancers rapidly evolving in a species- and/or lineage-specific manner (Villar et al., 2015).

A similar approach is to screen for regulatory elements within a single topologically associated domain (TAD). Genes and enhancers within a single TAD are typically isolated from those in neighboring TADs. To identify regulatory elements within an ~1-Mb TAD, one study first selected 174 regions in a TAD surrounding *POU5F1* (OCT4), a key pluripotency factor in stem cell self-renewal (Diao et al., 2016). These regions were selected based upon features that suggest gene regulatory function, such as enhancer-related chromatin marks, CTCF binding, and/or DHS. A saturating mutagenesis library with ~2,000 gRNAs with a mean of 11 gRNAs per element was used to find regulators of *POU5F1* expression. The screen was performed in *POUF51*-GFP stem cells so that GFP-negative cells could be sorted by FACS to identify altered *POU5F1* regulation. A recent extension of this work used pairs of gRNAs to cover ~2 Mb of sequence at the *POU5F1* locus, with 2-kb deletions achieving 20× coverage/overlap of each base (Diao et al., 2017). A similar tiling deletion screen approach was used to tile ~4,000 overlapping 1- to 2-kb deletions across 206 kb at the *HPRT1* locus and found a striking lack of distal regulatory elements capable of modulating *HPRT1* expression (Gasperini et al., 2017).

FACS-based approaches can be extended to virtually any expressed gene. For example, tiling gRNAs across regulatory elements (~40 kb of sequence) has been reported using GFP knockins at multiple genes (Rajagopal et al., 2016). Another recent FACS-based study examined murine enhancer elements near programmed cell death 1 (PD-1, Pcdh1), a key regulator of T cells and a cancer immunotherapy target, by interrogating nine distinct enhancer regions (Sen et al., 2016). This library included 1,754 enhancer-targeting gRNAs, 117 gRNAs targeting the PD-1 exonic sequence as positive controls, and 200 non-targeting gRNAs as negative controls. Using this approach, the authors identified an enhancer involved in T cell exhaustion and specific transcription factor binding sites responsible for the changes in T cell activity.

To increase the resolution of saturating mutagenesis, it is possible to screen the same region with multiple nucleases with different targeting capabilities (Canver et al., 2017a). For example, to identify regulatory elements for the *MYB* gene capable of modulating fetal hemoglobin, recent work combined SpCas9 (PAM: NGG) and the engineered PAM variant SpCas9-VQR (PAM: NGA). The combination library (with NGG and NGA PAMs) achieved a high level of saturation (3 bp median spacing between gRNAs), resulting in an ~2× increase in resolution from a single nuclease alone over the same region. The libraries were also designed using haplotype-derived variants instead of relying solely on the human reference genome, which was shown to reduce false negatives *in silico* due to mismatches between the gRNA and target sequence.

Noncoding Screens with CRISPR Effectors

Fusions of catalytically inactive CRISPR enzymes with other effector domains, such as those employed in CRISPRa and CRISPRi screens, have been effective at identifying regulatory

alter transcription factor binding motifs (or adjacent to transcription factor binding motifs) will display altered abundance when sorting for gene expression (Canver et al., 2015). By contrast, gRNAs that confer resistance to drugs or an aggressive *in vivo* metastatic phenotype will be more abundant (Sanjana et al., 2016; Chen et al., 2015; Shalem et al., 2014; Wang et al., 2014). Pooled screens have also been paired with single-cell RNA-seq to dissect genetic networks (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016).

regions. Using the dCas9-KRAB CRISPRi system, one study examined the regulatory architecture of regions around the *MYC* and *GATA1* genes, which are involved in cell proliferation, in a leukemia cell line. Decreased expression of these genes results in slower growth, which reduces the representation of the corresponding gRNAs over the growth period of the screen (14 population doublings). Using a pooled library of 98,000 gRNAs tiled across ~1.3 Mbp of sequence, this study found that a combination of chromosome conformation and enhancer marks like H3K27Ac can predict which regions harbor functional elements (Fulco et al., 2016). It is likely that dCas9 fused to other repressive domains like LSD1 (Kearns et al., 2015), HDAC (Kwon et al., 2017), MQ1 (Lei et al., 2017), or DNMT3A (Liu et al., 2016; Vojta et al., 2016) would also be successful in a similar pooled screening format. Another CRISPRi-based study coupled a noncoding screen to a single-cell RNA-seq readout to interrogate super-enhancers (Xie et al., 2017). The screen utilized dCas9-KRAB-expressing cells and a pooled, barcoded gRNA library targeting DHS regions within super-enhancers. In conjunction with gRNA identification, the transcriptome of each single cell was also acquired via the Drop-seq method (Macosko et al., 2015). Guide RNAs sparsely targeted 71 enhancers contained within 15 super-enhancers located in 7 distinct TADs. Interestingly, whereas most pooled screens are performed at low viral multiplicity to ensure single integrants per cell, this screen was performed at high viral multiplicity (an average of 3.2 gRNAs per cell), which is feasible given that the readout is not pooled (i.e., cells are barcoded individually). Transcriptomes of 12,444 single cells were obtained, which was estimated to be equivalent to 40,041 single-cell transcriptomes due to multiple gRNA integrants per cell. This analysis allowed for identification of functional enhancers and their effect on the transcriptome as well as identification of target genes. With multiple gRNAs delivered to each cell, this approach was also able to assess the effects of combinatorial targeting of several DHSs within a single super-enhancer.

Fewer studies have utilized CRISPRa in noncoding regions; this may be due to greater potential for false-positive results from inactive or poised enhancer elements. Using a dCas9-VP64 CRISPRa system, a recent study identified autoimmune-related enhancers of *CD69* and *IL2RA* by tiling ~100-kb regions around each gene (Simeonov et al., 2017). Another study utilized both CRISPRi and CRISPRa strategies over the same region (Klann et al., 2017). First, a library of 12,189 gRNAs was synthesized to target 433 DHSs in a 4-Mbp region flanking *HER2* and some non-hypersensitive sequences within the locus as a negative control. A screen was performed in the presence of dCas9-KRAB in A431 cells and used FACS to sort cells based on *HER2* expression levels; potential regulatory elements were identified based on gRNA enrichment in cells with low *HER2* expression. The same library and experimental design were then used in the presence of a dCas9 fusion with p300 in HEK293T cells. This novel CRISPRa approach takes advantage of p300, a histone acetyltransferase responsible for deposition of acetylation of histone H3 on lysine 27, a key biochemical modification found at active promoters and enhancers (Creyghton et al., 2010). The activation screen largely led to the identification of the same regulatory elements, with some differences between the sequences identified by the loss-of-function and gain-of-function techniques. The observed discrepancies were attributed to cell-type-specific enhancer activity, difficulties detecting active enhancers using activation approaches, and difficulties detecting inactive enhancers with repression approaches.

Lastly, targeted DNA methylation or enforced chromatin looping can be applied to enhancer/DHS evaluation. It has been shown that the fusion protein dCas9-DNMT3A can be used to catalyze DNA methylation of CpG island motifs in the targeted region to create stable, heritable silencing (Liu et al., 2016). This strategy is likely to be successful in a high-throughput pooled screening format. In addition, different dCas9 orthologs fused to inducible dimerzation components from the plant phytohormone S-(+)-abscisic acid (ABA) can be used to manipulate chromatin looping between different chromosomal regions, such as enhancers and promoters (Morgan et al., 2017).

Noncoding RNA Screens

lncRNAs, which are noncoding RNAs greater than 200 bp in length, have diverse functions in gene regulation and signaling, and as decoys and scaffolds for other RNAs (Wang and Chang, 2011). They are abundant in the human genome: GENCODE v26 has annotated 15,787 lncRNA genes and 27,720 lncRNA transcripts (Harrow et al., 2012). lncRNAs have been difficult to manipulate until recently. Recent work used the CRISPRa approach to target ~1 kb upstream of the first exon of 243 lncRNAs to identify lncRNAs regulators of AKT (Koirala et al., 2017). In another study, 671 human lncRNAs were targeted with 12,472 gRNA pairs for pooled deletion screening, which ultimately implicated 51 lncRNAs as playing a role in cancer cell growth (Zhu et al., 2016). CRISPRi has also been used to study lncRNAs. A 170,262 gRNA library was synthesized to target the TSS of 16,401 lncRNAs to identify hits playing a role in cell growth (Liu et al., 2017b). With CRISPRa, a 95,958 gRNA library was synthesized to target the TSS of 10,504 lncRNAs to identify key lncRNAs whose expression conferred resistance to a cancer therapy, including one lncRNA shown to regulate four nearby genes (Joung et al., 2017b). Similar to lncRNAs, pooled screens have been performed to interrogate microRNAs (Chen et al., 2015; Sanjana et al., 2014; Wallace et al., 2016). CRISPR-Cas9-mediated deletions have been shown to be an effective strategy to target individual microRNAs (Ho et al., 2015); therefore, it is conceivable that high-throughput pooled deletion screens may be an effective screening strategy for microRNAs as well. Alternatively, screens for noncoding RNAs could be performed using type VI effectors in a high-throughput format (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Smargon et al., 2017).

Transcription Factor Binding Site and Motif Targeted Screens

Transcription factor binding sites and motifs are key mediators of gene regulatory function, and many diseases, including cancer, can result from mutations in these regions (Cuykendall et al., 2017). Instead of targeting all binding sites around a specific gene of interest, it is possible to target binding sites genome-wide for a specific transcription factor. For example, one study interrogated p53-bound enhancer regions by targeting 685 genomic regions derived from overlapping ChIP-seq datasets, p53 consensus binding motifs, and enhancer annotations

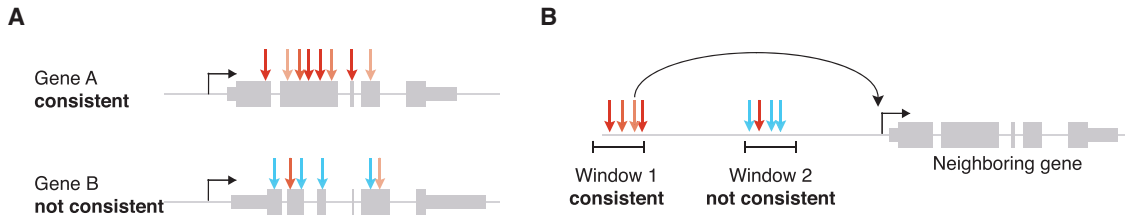

Figure 5. Consistent Enrichment/Depletion in Pooled CRISPR Screens
The analysis of pooled CRISPR screens is based on the abundance of gRNAs from next-generation sequencing of the genomically integrated gRNA cassette across samples. Enrichment and/or depletion scores are calculated by comparing the abundance after phenotypic selection with an earlier, pre-selection time point. Red arrows: gRNAs displaying enrichment across samples in the screen; blue arrows: gRNAs not showing changes in abundance across samples.
(A) In pooled screens targeting coding genes, the enrichment/dropout of multiple gRNAs targeting the same gene (consistent effects) enables identification of potential candidate genes. Genes where gRNAs are not consistently enriched/depleted are not chosen for further validation and analysis.
(B) In noncoding screens, candidate functional sequences are identified by enrichment or dropout of distinct gRNAs that target within a defined window.

(Korkmaz et al., 2016). A 1,116-gRNA tiling library was synthesized to evaluate the identified regions. A similar experiment was also performed for the transcription factor ERα by overlapping ChIP-seq datasets and ERα consensus binding motifs. In this case, enhancer RNA expression analyses by GRO-seq were utilized to find active enhancers. This analysis identified 73 candidate regions that were then interrogated using a 97-gRNA library. Both of these screens successfully identified functional enhancer sequences bound by the respective transcription factors.

CTCF is an insulator protein that has been shown to be important for genome organization, such as TAD formation. Disruption of CTCF binding sites can act as an oncogenic driver via enhancer hijacking and activation of neighboring oncogenes (Hnisz et al., 2016). Several tools exist for CTCF manipulation, such as nucleases for mutagenesis, sterically hindering CTCF binding (CRISPRi), methylation of CpG dinucleotides (dCas9-DNMT3A, dCas9-MQ1), and base editing approaches such as cytosine deaminase effectors (Hess et al., 2016; Lei et al., 2017; Liu et al., 2016; Ma et al., 2016). Combining these tools with a genome-wide CTCF library would allow rapid interrogation of the CTCF sites and TADs that are most relevant for particular diseases and phenotypes. Finally, it may be possible to perform a screen for DNA-binding partners for a given binding motif by using the new dCas9-APEX effector that biotinylates proteins near the gRNA target site (Myers et al., 2017). Similarly, a biotinylated dCas9 paired with pull-down for proteomics, 3C-seq, and RNA-seq can be used to characterize locus-specific chromatin-regulating protein complexes and long-range DNA interactions at specific motifs or enhancer elements (Liu et al., 2017c).

Analysis of Pooled CRISPR Screens

The initial step in the analysis of pooled CRISPR screens is converting sequencing data to enrichment and/or dropout ("depletion") scores for all gRNAs in the library. Typically, these scores are compared between pre- and post-selection samples (e.g., before and after drug selection to find gRNAs involved in drug resistance and/or sensitivity). Although most pooled CRISPR screens have compared gRNA representation between two samples, it is also possible to have a multi-dimensional phenotypic readout, such as FACS-based sorting of multiple populations to determine relevant effect size or single-cell full transcriptome readout (Adamson et al., 2016; Datlinger et al., 2017; Dixit et al., 2016; Jaitin et al., 2016; Xie et al., 2017). Many different tools have been developed for analysis of gene-targeted screens and offer such features as automated determination of positively/negatively selected genes, pathway analysis, quality control analysis, and data visualization (Jeong et al., 2017; Li et al., 2014, 2015; List et al., 2016; Winter et al., 2016).

For screens targeting protein-coding genes, the key task is to identify how many gRNAs that target the same gene are enriched/depleted and to identify those with consistent enrichment/depletion (Figure 5A). For example, genes with multiple distinct enriched/depleted gRNAs are ranked more highly than genes with only one enriched/depleted gRNA.

One outstanding question for the analysis of noncoding screens is how to accurately and reliably identify functionally relevant regions. Intuitively, the presence of colocalizing gRNAs that either enrich or drop out in a noncoding screen is suggestive of functional sequence (Figure 5B). One approach is to use a sliding window to identify regions with multiple enriched gRNAs (Diao et al., 2017; Fulco et al., 2016; Sanjana et al., 2016). This is similar to the consistent enrichment principle in screens targeting protein-coding genes, except that defining an appropriate window size introduces an additional free parameter. Other approaches to identify functional regions have included a hidden Markov model (HMM) (Canver et al., 2015, 2017a). Analytic methods for noncoding mutagenesis screens are still in their infancy, and there are abundant opportunities to develop more advanced methods.

After identification of screen hits, it is essential to validate the results of the screen. Validation of a gene-targeted screen typically employs independent gRNAs targeting the gene(s) of interest that are not found in the pooled library. For screens targeting noncoding regions (e.g., saturating mutagenesis), it can be challenging to find independent gRNAs if all nearby gRNAs were already included in the library or if there are not a sufficient number of targetable sites. In these cases, CRISPR nucleases with different PAM sequences can be utilized—either orthologs from other microbial species, or the same nuclease engineered to recognize an alternative PAM (Canver et al., 2017a). If there are many hits, or if validation in an array format is not feasible (e.g., with *in vivo* systems), it can be useful to construct a focused subpool library, with deep coverage of selected genes or regions from the initial pooled screen (Chen et al., 2015). The subpool library can be used to gain confidence in specific hits before proceeding to other validation assays.

Pitfalls of Pooled CRISPR Screening

On-target activity and off-target cleavage concerns are not unique to pooled screening experiments but rather represent concerns for all genome editing experiments. Many different predictive scores for on- and off-target activity have been developed to aid in gRNA choice to address these concerns (Tycko et al., 2016). In general, false negatives can occur in pooled screens when using gRNAs with low activity. To combat this problem, machine learning approaches have been applied to saturating mutagenesis screens in coding regions to identify gRNA characteristics that correlate with high activity (Doench et al., 2014, 2016). Similar approaches have also been deployed to derive design parameters for CRISPRi (Smith et al., 2016). In addition to utilizing scores effective at predicting on-target activity (Doench et al., 2014, 2016), other work has highlighted the importance of targeting nucleosome-free regions (Horlbeck et al., 2016) and methods for creating accessible chromatin to enhance on-target activity (Chen et al., 2017a). Others have found that nucleosome positioning has little or no effect in boosting activity (Chen et al., 2017c). Also, modifying the gRNA scaffold can improve nuclease-target binding and reduce the number of gRNAs required per target (Cross et al., 2016; Qi et al., 2013).

CRISPR targeting of genes or regions present at high copy number can result in apoptosis from excessive DSB induction and thus lead to false positives in dropout screens (Aguirre et al., 2016; Munoz et al., 2016). This can be especially problematic in the context of cancer, since many oncogenes (and noncoding regions near oncogenes) may be present at high copy number in cell lines or tumor samples (Santarius et al., 2010). These issues can be circumvented by either removing gRNAs with significant off-target potential at the library design stage or excluding them from the analysis after readout (Meyers et al., 2017).

Efforts to minimize false negatives and false positives are essential to ensure reproducibility. It can be challenging to directly assess reproducibility of high-throughput screens due to limited overlap of study questions and unique experimental conditions and designs. However, a number of gene-targeted studies have sought to identify essential genes in different cell types. These screens have, in general, identified similar sets of genes, although they also have found a smaller number of genes unique to a particular study and different relative essentiality ranks among genes identified in multiple studies (Hart et al., 2017; Rauscher et al., 2017a). It remains unclear whether these discrepancies are due to cell-type-specific effects, experimental artifacts, or differences in computational analysis of screen data. Notably, one group identified two enhancer elements using a saturating mutagenesis approach, which they re-identified using a tiling deletion-based approach (Diao et al., 2016, 2017). In addition, top hits from saturating mutagenesis of the *BCL11A* enhancer were re-validated in a subsequent saturating mutagenesis screen of the region (Canver et al., 2015, 2017a). Saturating mutagenesis using two Cas9 orthologs (SpCas9 and SpCas9-VQR) was performed separately on the same genomic locus. These experiments resulted in comparable functional maps, which offered internal reproducibility using independent CRISPR systems (Canver et al., 2017a). Finally, technical aspects of screen design can also factor into reproducibility. Independent screen replicates using the same library and cell line are usually highly reproducible, with a notable exception being CRISPR screens using homologous recombination instead of lentivirus for library delivery (Rajagopal et al., 2016). Taken together, initial studies have suggested a high degree of reproducibility of pooled CRISPR screens, but systematic investigation of this issue is still needed.

Consideration of off-target activity is also important in large-scale screens: given that libraries are composed of many gRNAs, it is necessary to have some assurances that targeting is specific to the intended region of the genome. One advantage of pooled screens is that the downstream analysis includes some measure of consistency between multiple gRNAs targeting the same gene or noncoding region (Figures 5A and 5B). Thus, the off-target effects of a single gRNA are less problematic in the context of screens than in, for example, therapeutic development. Early work noted that mismatches in certain positions are less likely to induce off-target activity (Hsu et al., 2013), and more recent work has used large-scale screen datasets to develop rules to predict off-target activity such as the cutting-frequency determination (CFD) score (Doench et al., 2016). Another approach to reducing off-target activity is to use nucleases engineered with reduced off-target potential such as HypaCas9, SpCas9-HF1, and eSpCas9 (Chen et al., 2017b; Kleinstiver et al., 2016; Slaymaker et al., 2016). A final consideration concerns the use of non-targeting gRNAs as negative controls. It has been suggested that gRNAs that target nonessential genes are a superior negative control for pooled screens because they better account for Cas9 toxicity and DSB induction (Hart et al., 2017; Morgens et al., 2017).

Conclusions

Pooled CRISPR screens are a powerful tool to perform high-throughput genetic and epigenetic studies. Forward genetic screens using genome engineering enable investigators to establish causal links between noncoding regions and disease-relevant phenotypes. Given that tools to establish direct links between noncoding sequences in their native context and function have not been available until recently, there is a growing interest in developing and applying high-throughput CRISPR-based screens for these purposes. A growing database of CRISPR screens highlights the importance of this data and offers a rich resource for researchers (Rauscher et al., 2017b). We expect that pooled screens in noncoding regions will evolve rapidly as CRISPR technology continues to improve. It is possible that in the not-too-distant future, extensive mapping of noncoding function catalyzed by these tools will provide a clear picture of how much of the human genome is truly functional.

AUTHOR CONTRIBUTIONS

A.M. and M.C.C. reviewed the literature. A.M., M.C.C., and N.E.S. wrote the manuscript.

ACKNOWLEDGMENTS

M.C.C. is supported by a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Award (F30-DK103359). N.E.S. is supported by the National Institutes of Health through the National Human Genome

Research Institute (R00-HG008171), by the Kimmel Foundation, and by the Melanoma Research Alliance. N.E.S. is listed as an inventor on patent applications related to this work.

REFERENCES

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1882.e21.

Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.-Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer Discov. *6*, 914–929.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. *34*, 933–941.

Barrangou, R., and Gersbach, C.A. (2017). Expanding the CRISPR Toolbox: Targeting RNA with Cas13b. Mol. Cell *65*, 582–584.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A., and Banfield, J.F. (2017). New CRISPR-Cas systems from uncultivated microbes. Nature *542*, 237–241.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature *527*, 192–197.

Canver, M.C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E.N., Needleman, A.J., Galactéros, F., Brugnara, C., Kutlar, A., et al. (2017a). Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. Nat. Genet. *49*, 625–634.

Canver, M.C., Haeussler, M., Bauer, D.E., Orkin, S.H., Sanjana, N.E., Shalem, O., Yuan, G.-C., Zhang, F., Concordet, J.-P., and Pinello, L. (2017b). Integrated design, execution, and analysis of arrayed and pooled CRISPR genome editing experiments. bioRxiv. http://dx.doi.org/10.1101/125245.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell *160*, 1246–1260.

Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y., and Davis, G.D. (2017a). Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat. Commun. *8*, 14958.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Harrington, L.B., Sternberg, S.H., Joung, K.J., Yildiz, A., and Doudna, J.A. (2017b). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. bioRxiv. http://dx.doi.org/10.1101/160036.

Chen, Y., Zeng, S., Hu, R., Wang, X., Huang, W., Liu, J., Wang, L., Liu, G., Cao, Y., and Zhang, Y. (2017c). Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish. PLoS ONE *12*, e0182528.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. *23*, 1163–1171.

Chow, R.D., Guzman, C.D., Wang, G., Schmidt, F., Youngblood, M.W., Ye, L., Errami, Y., Dong, M.B., Martinez, M.A., Zhang, S., et al. (2017). AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma. Nat. Neurosci. Published online August 14, 2017. http://dx.doi.org/10.1038/nn.4620.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Cross, B.C.S., Lawo, S., Archer, C.R., Hunt, J.R., Yarker, J.L., Riccombeni, A., Little, A.S., McCarthy, N.J., and Moore, J.D. (2016). Increasing the performance of pooled CRISPR-Cas9 drop-out screening. Sci. Rep. *6*, 31782.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Cuykendall, T.N., Rubin, M.A., and Khurana, E. (2017). Non-coding genetic variation in cancer. Curr. Opin. Syst. Biol. *1*, 9–15.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301.

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A.Y., Dixon, J., Maliskova, L., Guan, K.L., Shen, Y., and Ren, B. (2016). A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. Genome Res. *26*, 397–405.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat. Methods *14*, 629–635.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. *32*, 1262–1267.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Donovan, K.F., Hegde, M., Sullender, M., Vaimberg, E.W., Johannessen, C.M., Root, D.E., and Doench, J.G. (2017). Creation of novel protein variants with CRISPR/Cas9-mediated mutagenesis: Turning a screening by-product into a discovery tool. PLoS ONE *12*, e0170445.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270–273.

Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.

Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. Nature *513*, 120–123.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science *354*, 769–773.

Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. Nat. Biotechnol. *35*, 789–792.

Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. Am. J. Hum. Genet. *101*, 192–205.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Graur, D. (2017). An upper limit on the functional fraction of the human genome. Genome Biol. Evol. *9*, 1880–1885.

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. *23*, 2126–2135.

Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat. Biotechnol. *35*, 463–474.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 (Bethesda) *7*, 2719–2727.

Hess, G.T., Frésard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Methods *13*, 1036–1042.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454–1458.

Ho, T.T., Zhou, N., Huang, J., Koirala, P., Xu, M., Fung, R., Wu, F., and Mo, Y.Y. (2015). Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. Nucleic Acids Res. *43*, e17.

Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R., and Weissman, J.S. (2016). Nucleosomes impede Cas9 access to DNA in vivo and in vitro. eLife *5*, e12677.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell *167*, 1883–1896.e15.

Jeong, H.H., Kim, S.Y., Rousseaux, M.W.C., Zoghbi, H.Y., and Liu, Z. (2017). CRISPRcloud: A secure cloud-based pipeline for CRISPR pooled screen deconvolution. Bioinformatics. Published online May 24, 2017. http://dx.doi.org/10.1093/bioinformatics/btx335.

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

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017a). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.

Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., et al. (2017b). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature *548*, 343–346.

Karvelis, T., Gasiunas, G., and Siksnys, V. (2017). Harnessing the natural diversity and in vitro evolution of Cas9 to expand the genome editing toolbox. Curr. Opin. Microbiol. *37*, 88–94.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods *12*, 401–403.

Kim, D., Lim, K., Kim, S.-T., Yoon, S.H., Kim, K., Ryu, S.-M., and Kim, J.-S. (2017a). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat. Biotechnol. *35*, 475–480.

Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S., et al. (2017b). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat. Commun. *8*, 14500.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017c). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. *35*, 371–376.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015a). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015b). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. *33*, 1293–1298.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, Mdel.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267–273.

Koirala, P., Huang, J., Ho, T.-T., Wu, F., Ding, X., and Mo, Y.-Y. (2017). LncRNA AK023948 is a positive regulator of AKT. Nat. Commun. *8*, 14422.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Konermann, S., Brigham, M.D., Trevino, A., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472–476.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78.

Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. *34*, 192–198.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., and Adli, M. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat. Methods *14*, 710–712.

Kwon, D.Y., Zhao, Y.-T., Lamonica, J.M., and Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. Nat. Commun. *8*, 15315.

Lei, Y., Zhang, X., Su, J., Jeong, M., Gundry, M.C., Huang, Y.-H., Zhou, Y., Li, W., and Goodell, M.A. (2017). Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat. Commun. *8*, 16026.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Li, W., Köster, J., Xu, H., Chen, C.-H., Xiao, T., Liu, J.S., Brown, M., and Liu, X.S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biol. *16*, 281.

List, M., Schmidt, S., Christiansen, H., Rehmsmeier, M., Tan, Q., Mollenhauer, J., and Baumbach, J. (2016). Comprehensive analysis of high-throughput screens with HiTSeekR. Nucleic Acids Res. *44*, 6639–6648.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233–247.e17.

Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., and Wang, Y. (2017a). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. Mol. Cell *65*, 310–322.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017b). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science *355*, aah7111.

Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., et al. (2017c). In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell *170*, 1028–1043.e19.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods *13*, 1029–1035.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods *10*, 977–979.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013a). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013b). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Manguso, R.T., Pope, H.W., Zimmer, M.D., Brown, F.D., Yates, K.B., Miller, B.C., Collins, N.B., Bi, K., LaFleur, M.W., Juneja, V.R., et al. (2017). In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. Nature *547*, 413–418.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Meier, J.A., Zhang, F., and Sanjana, N.E. (2017). GUIDES: sgRNA design for loss-of-function screens. Nat. Methods *14*, 831–832.

Meyers, R.M., Bryan, J.G., Mcfarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. bioRxiv. http://dx.doi.org/10.1101/160861.

Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.-F., et al. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat. Commun. *8*, 15993.

Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E., et al. (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. Nat. Commun. *8*, 15178.

Mou, H., Smith, J.L., Peng, L., Yin, H., Moore, J., Zhang, X.-O., Song, C.-Q., Sheel, A., Wu, Q., Ozata, D.M., et al. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol. *18*, 108.

Munoz, D.M., Cassiani, P.J., Li, L., Billy, E., Korn, J.M., Jones, M.D., Golji, J., Ruddy, D.A., Yu, K., McAllister, G., et al. (2016). CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. Cancer Discov. *6*, 900–913.

Myers, S.A., Wright, J., Zhang, F., and Carr, S.A. (2017). CRISPR/Cas9-APEX-mediated proximity labeling enables discovery of proteins associated with a predefined genomic locus in living cells. BioRxiv. http://dx.doi.org/10.1101/159517.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, 553–563.

Ohno, S. (1972). An argument for the genetic simplicity of man and other mammals. J. Hum. Evol. *1*, 651–662.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455.

Pott, S. (2017). Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. eLife *6*, e23203.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Radzisheuskaya, A., Shlyueva, D., Müller, I., and Helin, K. (2016). Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. Nucleic Acids Res. *44*, e141.

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J.M., Gifford, D.K., and Sherwood, R.I. (2016). High-throughput mapping of regulatory DNA. Nat. Biotechnol. *34*, 167–174.

Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z., and Shendure, J. (2017). Massively multiplex single-cell Hi-C. Nat. Methods *14*, 263–266.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191.

Rands, C.M., Meader, S., Ponting, C.P., and Lunter, G. (2014). 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. PLoS Genet. *10*, e1004525.

Rauscher, B., Heigwer, F., Henkel, L., Hielscher, T., and Boutros, M. (2017a). A map of genetic interactions in cancer cells. bioRxiv. http://dx.doi.org/10.1101/120964.

Rauscher, B., Heigwer, F., Breinig, M., Winter, J., and Boutros, M. (2017b). GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens. Nucleic Acids Res. *45* (D1), D679–D686.

Rosenbluh, J., Xu, H., Harrington, W., Gill, S., Wang, X., Vazquez, F., Root, D.E., Tsherniak, A., and Hahn, W.C. (2017). Complementary information

derived from CRISPR Cas9 mediated gene deletion and suppression. Nat. Commun. *8*, 15403.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat. Biotechnol. *33*, 1165–1172.

Sanjana, N.E. (2017). Genome-scale CRISPR pooled screens. Anal. Biochem. *532*, 95–99.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F. (2016). High-resolution interrogation of functional elements in the noncoding genome. Science *353*, 1545–1549.

Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. Nat. Rev. Cancer *10*, 59–64.

Sen, D.R., Kaminski, J., Barnitz, R.A., Kurachi, M., Gerdemann, U., Yates, K.B., Tsao, H.-W., Godec, J., LaFleur, M.W., Brown, F.D., et al. (2016). The epigenetic landscape of T cell exhaustion. Science *354*, 1165–1169.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. *16*, 299–311.

Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat. Methods *14*, 573–576.

Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat. Biotechnol. *33*, 661–667.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods *11*, 817–820.

Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., et al. (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol. Cell *65*, 618–630.e7.

Smith, J.D., Suresh, S., Schlecht, U., Wu, M., Wagih, O., Peltz, G., Davis, R.W., Steinmetz, L.M., Parts, L., and St Onge, R.P. (2016). Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. Genome Biol. *17*, 45.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Thakore, P.I., Black, J.B., Hilton, I.B., and Gersbach, C.A. (2016). Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods *13*, 127–137.

Tycko, J., Myer, V.E., and Hsu, P.D. (2016). Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. Mol. Cell *63*, 355–370.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Vitak, S.A., Torkenczy, K.A., Rosenkrantz, J.L., Fields, A.J., Christiansen, L., Wong, M.H., Carbone, L., Steemers, F.J., and Adey, A. (2017). Sequencing thousands of single-cell genomes with combinatorial indexing. Nat. Methods *14*, 302–308.

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M., and Zoldoš, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. *44*, 5615–5628.

Wallace, J., Hu, R., Mosbruger, T.L., Dahlem, T.J., Stephens, W.Z., Rao, D.S., Round, J.L., and O'Connell, R.M. (2016). Genome-Wide CRISPR-Cas9 Screen Identifies MicroRNAs That Regulate Myeloid Leukemia Cell Growth. PLoS ONE *11*, e0153689.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Wang, T., Lander, E.S., and Sabatini, D.M. (2016). Viral packaging and cell culture for CRISPR-based screens. Cold Spring Harb. Protoc. *2016*, t090811.

Wang, G., Chow, R., Ye, L., Guzman, C., Dai, X., Dong, M., Zhang, F., Sharp, P., Platt, R., and Chen, S. (2017). Pooled AAV-CRISPR screen with targeted amplicon sequencing. bioRxiv. http://dx.doi.org/10.1101/153643.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Winter, J., Breinig, M., Heigwer, F., Brügemann, D., Leible, S., Pelz, O., Zhan, T., and Boutros, M. (2016). caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. Bioinformatics *32*, 632–634.

Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M., Perli, S.D., Kazer, S.W., Gaillard, A., Hermann, M., et al. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc. Natl. Acad. Sci. USA *113*, 2544–2549.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48–60.e11.

Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G.C. (2017). Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. Mol. Cell *66*, 285–299.e5.

Xu, C., Qi, X., Du, X., Zou, H., Gao, F., Feng, T., Lu, H., Li, S., An, X., Zhang, L., et al. (2017). *piggyBac* mediates efficient in vivo CRISPR library screening for tumorigenesis in mice. Proc. Natl. Acad. Sci. USA *114*, 722–727.

Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G., et al. (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature *514*, 380–384.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., et al. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat. Biotechnol. *35*, 31–34.

Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., et al. (2016). Genome-scale deletion screening of human long noncoding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nat. Biotechnol. *34*, 1279–1286.