Published in final edited form as:

Curr Opin Genet Dev. 2018 October ; 52: 70–76. doi:10.1016/j.gde.2018.06.001.

Decoding the noncoding genome via large-scale CRISPR screens

Abhijit Shukla¹ and Danwei Huangfu^{1,*}

¹Sloan Kettering Institute, 1275 York Avenue, New York, New York 10065, USA

Abstract

Large portions of the human genome harbor functional noncoding elements, which can regulate a variety of biological processes and have important implications for disease risk and therapeutic outcomes. However, assigning specific functions to noncoding sequences remains a major challenge. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated protein (Cas) systems have emerged as a powerful approach for targeted genome and epigenome perturbation. CRISPR systems are now harnessed for high-throughput screening of the noncoding genome to uncover functional regulatory elements and to define their precise functions with superior speed. Here, we summarize the various tools developed for such screens in mammalian systems, discuss screening methods and technical considerations, highlight screens that are already transforming our understanding of gene regulation and disease mechanisms, consider the impact of such discoveries on the development of new therapeutics, and provide our viewpoint on the challenges for future development of the field.

Introduction

Precise spatial and temporal regulation of gene expression is crucial for embryonic development, tissue homeostasis and regeneration. Less than 2% of the human genome harbors protein-coding genes, with the rest encompassing noncoding sequences [1]. Greater than 90% of disease and trait-associated variants from genome-wide association studies (GWAS) are found within the noncoding genome, emphasizing their importance to human health [2]. However, the precise function of most noncoding sequences is yet to be elucidated [3, 4]. Consortium efforts like the Encyclopedia of DNA Elements (ENCODE) and the National Institutes of Health (NIH) Roadmap Epigenomics Project have produced rich datasets about the chromatin epigenetic landscape (such as DNA methylation, histone acetylation/methylation), transcription factor binding and chromatin accessibility [5]. These data are already being used to predict functional elements such as promoters, enhancers and functional noncoding RNAs, but the majority of these predictions are yet to be experimentally validated. Indeed, current annotation methods typically assign putative

*Correspondence to: huangfud@mskcc.org(D.H.).

The authors declare no conflict of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

enhancers to their nearest neighboring genes, which could lead to high false discovery rates [6]. Therefore, there is a pressing need for highthroughput approaches to assign causality through unbiased perturbation followed by functional evaluation in the context of development and disease.

Episomal reporter based large-scale assays such as massively parallel reporter assays (MPRAs) and self-transcribing active regulatory region sequencing (STARR-seq) have been employed to simultaneously evaluate the activity of thousands of potential expression-modulating DNA sequence variants [7,8]. However, these methods are thus far limited by the maximum number (typically <100,000) of oligonucleotides and the size (typically <200 bp) of the fragments that can be tested, and perhaps most importantly, the lack of complete native features. Recently, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated protein (Cas) systems have attracted great attention as they support high-throughput perturbation of noncoding element in its native chromatin context [9**]. In this review, we will discuss various aspects of CRISPR-based noncoding genetic screens in terms of choice of perturbation and screening methods, as well as current challenges and potential opportunities.

Cas9 and dCas9-based CRISPR perturbation methods

The CRISPR/Cas9 nuclease system is a type of prokaryotic adaptive immune system that provides resistance against bacteriophages [10]. Cas9 is an RNA-guided DNA endonuclease that induces double strand breaks (DSBs). This system has been repurposed for eukaryotic genome editing by designing engineered single guide RNA (sgRNA) that targets the heterologous Cas9 protein to a specific genomic locus with the “protospacer adjacent motif” (PAM) sequence and a preceding 20-nt target sequence matching the spacer sequence in the sgRNA [11-13]. Most CRISPR screens so far have used *S. pyogenes* Cas9 (SpCas9), which recognizes the NGG PAM sequence. Future screens are likely to also make use of the expanding list of Cas9 variants and CRISPR/Cas systems [14].

The Cas9 protein has been modified to generate a nuclease-deficient version (dCas9) that does not induce DNA cleavage but, rather, allows targeted chromatin modification [15]. By fusing dCas9 to transcriptional repressors or activators, CRISPR interference (CRISPRi) and activation (CRISPRa) methods have been developed to specifically suppress or enhance target gene expression [16**]. For CRISPRi, dCas9 can be fused with a number of repressive domains including LSD1, HDAC, MQ1, and DNMT3A [16**]. CRISPRi screens have primarily used dCas9 fused to the Krüppel-associated box (KRAB) domain, which induces histone modifications associated with heterochromatin [17]. For CRISPRa, dCas9 can be linked to single or multiple activator domains (e.g. p300, VP64) via direct fusion or recruitment through protein (e.g. SunTag) or RNA (e.g. MS2 RNA hairpin) scaffolds [18]. For instance, the dCas9 SunTag system recruits multiple copies of the transcriptional activator VP64, whereas the SAM (Synergistic Activation Mediator) system utilizes sgRNAs with the MS2 RNA aptamer to recruit MS2 RNA coat protein (MCP) fused with p65 and Heat shock factor 1 (HSF1) transactivation domains. Most CRISPRa screens so far have used the dCas9-VP64 system, and some have used dCas9-p300 and SAM systems (Table 1).

Screening approaches

Following the success of early studies using individual sgRNAs to interrogate enhancer-like areas such as transcription factor binding sites [19-21], larger scale screens have been performed using Cas9 and pooled sgRNA libraries to saturate a chosen noncoding region with indel (insertion/deletion) mutations (Table 1). Similarly, CRISPRi and CRISPRa approaches are being used to discover hidden regulatory elements. Genetic screens can be performed in pooled or arrayed formats.

In pooled screens, cells are infected with a lentiviral library of CRISPR perturbation reagents *en masse*, with the objective of integrating one lentivirus carrying one sgRNA per cell (Figure 1). After the screen, the relative abundance of individual sgRNAs, identified by the unique spacer sequences, can be quantified by next-generation sequencing followed by data deconvolution. A number of pooled screens have been performed to interrogate cell growth or survival phenotypes including drug resistance. Typically, the infected cells are subjected to a positive or negative selection method designed to enrich or deplete the sgRNAs of interest, respectively. Apart from cell growth or viability, gene expression levels can also be used as readouts. Fluorescence-activated cell sorting (FACS) has been used to identify sgRNAs that cause changes in target gene expression. The fluorescent signal can be derived from a cell line with a knock-in or transgenic fluorescent reporter, or by a fluorescently labeled antibody detecting a protein of interest [22**, 23**, 24-27]. The pooled format is easily scalable and suitable for screens using tiling sgRNA libraries to achieve near saturation perturbation within a noncoding region of interest. A major limitation is that the pooled format typically does not support readouts beyond cell growth, survival, or individual marker gene expression. To overcome this limitation, emerging methods combining pooled screens with single-cell RNA sequencing (scRNA-seq) [28-31] have recently been applied to interrogate noncoding sequences [32**].

As an alternative to pooled screens, arrayed screens have the potential to support high-content readouts such as cell morphology, function, protein localization, and comprehensive transcriptomic analysis through RNA-seq. Arrayed library screens are typically performed in 96- or 384-well plates, with each well containing a unique CRISPR reagent. Small-scale arrayed screens have been performed through the transfection of CRISPR plasmids or synthetic CRISPR reagents targeting protein-coding genes [33, 34]. Genome-scale synthetic sgRNA libraries targeting protein-coding genes are now becoming available from companies such as Synthego. Due to scalability limitations, the arrayed format is yet to be applied to high-throughput screening of noncoding sequences.

CRISPR/Cas9-based deletion screens

Proximal *cis*-regulatory elements of a gene can be identified by interrogating nearby sequences with Cas9-induced small indel mutations. Typically, a tiling sgRNA library is designed to target nearly all possible targeting sites in the region of interest. In an excellent example, Feng Zhang and colleagues first performed a genome-scale CRISPR screen for protein-coding gene mutations that confer resistance to the BRAF inhibitor vemurafenib in human melanoma cells [35]. A library of 18,315 sgRNAs was then designed to target >700

kb sequences surrounding three genes (*NF1*, *NF2* and *CUL3*) identified from the first study [35], to determine *cis*-regulatory elements that mediate vemurafenib resistance [36]. In another study, Richard Sherwood and colleagues used mouse embryonic stem cells carrying a knock-in green fluorescence protein (GFP) reporter in the *Tdgfla* locus, and screened 3,908 sgRNAs tiled across its ~40-kb proximal regulatory region [23**]. This study found several *cis*-regulatory elements that do not overlap with known enhancer features based on genomic profiling, emphasizing the need for unbiased CRISPR screens to reveal hidden regulatory elements.

The two main challenges associated with the screening of noncoding sequences are how to increase the resolution and the coverage of the screens. The resolution of the mutagenesis screen depends on the density of available PAM sequences, which can be underrepresented in some genomic areas. To overcome this challenge, one strategy is to use multiple Cas9 variants with different PAM sequences to increase the density of target sites [37*]. A second, and perhaps most important, challenge is the sheer size of the noncoding genome. Although it is possible to use a tiling sgRNA library to mutate a small noncoding region (typically less than 1-Mb) to near saturation, to expand this strategy to distal enhancers or larger genomic regions would necessitate an unrealistically large CRISPR library. One strategy to overcome this challenge is to introduce larger deletions using paired sgRNAs [26,38]. For example, Bing Ren and colleagues interrogated a 2-Mb region surrounding the *POU5F1* (*OCT4*) locus using 11,570 sgRNA pairs in human embryonic stem cells and identified 45 *cis*-regulatory elements [26]. A second approach, and the one most widely used, involves pre-identifying putative regulatory elements within a large genomic region based on genomic data and computational analysis [23**,24,39**]. For example, Reuven Agami and colleagues designed 1,116 sgRNAs to target 685 p53 binding sites selected based on p53 ChIP-seq and ENCODE datasets, and identified key p53 targets that mediate oncogenic RAS induced cellular senescence [39**].

CRISPRi and CRISPRa screens

In addition to CRISPR/Cas9-based deletion screens, CRISPRi and CRISPRa screens are emerging as effective tools for identifying regulatory regions (Table 1). In the first demonstration of CRISPRi screens for functional noncoding elements, Jesse Engreitz, Eric Lander and colleagues used the dCas9-KRAB system and a pooled library of 98,000 sgRNAs tiled across a 1.29-Mb region surrounding the *GATA1* and *MYC* genes. The study successfully identified nine distal enhancers that affect the proliferation of K562 human erythroleukemia cells through regulating the expression of *GATA1* and *MYC* [40**]. More recently Gary Hon and colleagues took CRISPRi screens to the next level, incorporating a barcoded sgRNA library and scRNA-seq readout to interrogate super-enhancers in K562 cells [32**].

In parallel to CRISPRi, the development of CRISPRa tools have not only renewed interest in gene over-expression screens, but have also enabled new ways of interrogating *cis*-regulatory elements. In a recent study, Alexander Marson and colleagues used dCas9-VP64 and sgRNAs covering ~100-kb sequences surrounding *CD69* and *IL2RA* and identified several CRISPRa- responsive elements (CaREs) [27**]. One of the CaREs has been

implicated in autoimmunity risk, which supports the biological relevance of CaREs though their roles in *cis*-regulation are yet to be fully elucidated [27**].

CRISPR/Cas9, CRISPRi and CRISPRa are being rapidly adopted as complementary screening methods. For instance, to discover novel functions of long noncoding RNAs (lncRNAs), screens have been performed using CRISPR/Cas9 with paired sgRNAs, dCas9-KRAB and dCas9-VP64 [41**,42*,43]. In one notable example, Charles Gersbach and colleagues compared dCas9-KRAB and dCas9-p300 screens conducted in multiple cell lines [22**]. They found both shared and cell type-specific enhancer activities; CRISPRa screen may miss active regulatory elements, whereas CRISPRi may miss repressed regulatory elements. They further compared a CRISPRa screen performed in HEK293T (with low *HER2* expression), to a CRISPRi screen performed in A431 cells (with moderate *HER2* expression) on *cis*-regulation of *HER2*, and found largely correlative results from the two screens, although there are also some differences. This study also found more complete identification of functional *cis*-regulatory elements in screens using dCas9-KRAB compared to CRISPR/Cas9. With the rapid improvement of Cas9- and dCas9-based screening approaches, the selection of the most suitable method(s) for specific screening purposes can benefit from more comparative studies in future.

Challenges and future perspectives

Since the first implementation of CRISPR/Cas9 in mammalian cells in 2013 [11-13], the quickly expanding CRISPR toolbox is not only transforming biomedical research in general with studies into the protein-coding region, but is also now providing further insight through interrogation of the noncoding genome for both basic research and therapeutic application. Indeed, researchers have already conducted screens to explore disease-relevant phenotypes beyond cancer drug resistance and oncogene-induced cellular senescence. For example, Daniel Bauer and colleagues designed 702 sgRNAs to target three DNase I hypersensitive sites (DHSs) within a 12-kb region at the *BCL11A* locus previously identified from GWAS as an ameliorating factor in β -thalassemia and sickle-cell anemia [44]. The study identified a specific enhancer element that can be exploited to reactivate fetal hemoglobin in adult patients to ameliorate β -globin deficiencies [25]. Another study used 1,754 sgRNAs to target putative enhancer regions (based on ATAC-seq data) of *PD-1* (*PDCDI*), a tumor immunotherapy target [45]. The *PD-1* enhancer identified from the screen may be explored to reverse T cell exhaustion in cancers. The success of these screens is a major step towards identifying therapeutic targets for pathological conditions.

There are several areas in which CRISPR screens may be further improved, including increasing target specificity and expanding target range, designing better CRISPR libraries such as paired sgRNA libraries for more efficient creation of large deletions and for genetic interaction studies, and incorporating powerful technologies such as single cell sequencing, CyTOF and cellular barcoding. Here we highlight two particular areas. The first involves the challenge of screening the noncoding genome due to its vast size. There is no available method for unbiased, saturated perturbation of the entire noncoding genome. Current methods allow either saturating a specific locus with CRISPR perturbations, or targeting pre-identified areas in a large genomic region based on specific genomic features. The latter

could be facilitated by improved genomic assays such as the recently developed HiChIP method [46] along with computational innovations that better predict functional noncoding elements.

The second area of importance lies in finding suitable biological contexts to address a broad range of questions from fundamental principles of biology to human health. For studies of development or disease-relevant phenotypes, it is important to conduct screens on primary cells, in particular human cells. However, thus far the majority of screens are conducted in cancer cell lines, with relatively few screens performed in primary human and mouse cells [34,47]. In this regard, we are particularly excited about the potential of *in vivo* screens and screens utilizing human pluripotent stem cells (PSCs). There are already examples of screens targeting protein-coding or noncoding regions in cultured cells, which were then transplanted into animal models [48,49]. In a recent study, Sidi Chen and colleagues introduced CRISPR libraries directly into the mouse brain to screen for glioblastoma suppressor genes [50]. Future studies may extend this approach to the noncoding genome and additional tissue types, although *in vivo* CRISPR delivery could pose a significant hurdle for many organ systems. Complementing *in vivo* animal studies, one may utilize the differentiation capacity of human PSCs to generate disease-relevant cell types that are otherwise difficult to obtain, and even create three-dimensional “mini organs” known as organoids. The small number of screens thus far performed in mouse and human PSCs have focused on the self-renewing stage [23**,24,26,42*]. We foresee that future screens will utilize human PSC lines such as those engineered to inducibly express Cas9 or dCas9-KRAB [52, 53], and make use of their differentiation capacity for studies of both the protein-coding and noncoding genomes. Ultimately through concerted efforts in a variety of *in vivo* and *in vitro* systems, CRISPR screens will define novel gene regulatory mechanisms and hence could pave a path towards the development of more effective therapeutics.

Acknowledgements

The work in the Huangfu laboratory is supported by NIH/NIDDK (R01DK096239), New York State Stem Cell Science (NYSTEM C029567 and C029156), Tri-Institutional Stem Cell Initiative (#2016-004, #2016-032), JDRF (3-SRA-2017-364-S-B), and the MSKCC Cancer Center Support Grant (P30CA008748). We thank Julian Pulecio, Renhe Luo and Chew-Li Soh for thoughtful comments on the manuscript. Susan Weil and Chew-Li Soh assisted with the graphics.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted

* of special interest

** of outstanding interest

1. Cheng Y, Ma Z, Kim BH, Wu W, Cayting P, Boyle AP, Sundaram V, Xing X, Dogan N, Li J, Euskirchen G et al.: Principles of regulatory information conservation between mouse and human. *Nature* (2014) 515(7527):371–375. [PubMed: 25409826]
2. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A et al.: Systematic localization of common disease-associated variation in regulatory DNA. *Science* (2012) 337(6099):1190–1195. [PubMed: 22955828]

3. Consortium EP: An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489(7414):57–74. [PubMed: 22955616]
4. Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E: On the immortality of television sets: “Function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol* (2013) 5(3):578–590. [PubMed: 23431001]
5. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ et al.: Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518(7539):317–330. [PubMed: 25693563]
6. Sanyal A, Lajoie BR, Jain G, Dekker J: The long-range interaction landscape of gene promoters. *Nature* (2012) 489(7414):109–113. [PubMed: 22955621]
7. Dailey L: High throughput technologies for the functional discovery of mammalian enhancers: New approaches for understanding transcriptional regulatory network dynamics. *Genomics* (2015) 106(3):151–158. [PubMed: 26072436]
8. White MA: Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics* (2015) 106(3):165–170. [PubMed: 26072432]
9. Montalbano A, Canver MC, Sanjana NE: High-throughput approaches to pinpoint function within the noncoding genome. *Mol Cell* (2017) 68(1):44–59. [PubMed: 28985510] ** A comprehensive review on applications of CRISPR-Cas9 technology in genetic screens with a broad discussion over the design and analysis of noncoding screens.
10. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: CRISPR provides acquired resistance against viruses in prokaryotes. *Science* (2007) 315(5819):1709–1712. [PubMed: 17379808]
11. Jinek M, East A, Cheng A, Ling S, Ma E, Doudna JA: RNA-programmed genome editing in human cells. *Elite* (2013), 2: e00471.
12. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: Multiplex genome engineering using CRISPR/Cas systems. *Science* (2013) 339(6121):819–823. [PubMed: 23287718]
13. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: RNA-guided human genome engineering via Cas9. *Science* (2013) 339(6121):823–826. [PubMed: 23287722]
14. Murugan K, Babu K, Sundaresan R, Rajan R, Sashital DG: The revolution continues: Newly discovered systems expand the CRISPR-Cas toolkit. *Mol Cell* (2017) 68(1):15–25. [PubMed: 28985502]
15. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA: Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* (2013) 152(5):1173–1183. [PubMed: 23452860]
16. Pulecio J, Verma N, Mejia-Ramirez E, Huangfu D, Raya A: CRISPR/Cas9-based engineering of the epigenome. *Cell Stem Cell* (2017) 21(4):431–447. [PubMed: 28985525] ** This comprehensive review summarises the latest developments in the field of CRISPR-Cas9 based epigenome editing. Along with discussing technical aspects of epigenome engineering by CRISPR-Cas9, this review also provides a bird’s eye view on the applications of CRISPR-Cas9 in stem cell biology and regenerative medicine.
17. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim et al.: CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* (2013) 154(2):442–451. [PubMed: 23849981]
18. Shalem O, Sanjana NE, Zhang F: High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* (2015) 16(5):299–311. [PubMed: 25854182]
19. Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, Bradner JE, Young RA: Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* (2015) 58(2):362–370. [PubMed: 25801169]
20. Hewitt KJ, Kim DH, Devadas P, Prathibha R, Zuo C, Sanalkumar R, Johnson KD, Kang YA, Kim JS, Dewey CN, Keles S et al.: Hematopoietic signaling mechanism revealed from a stem/progenitor cell cistrome. *Mol Cell* (2015) 59(1):62–74. [PubMed: 26073540]

21. Puente XS, Bea S, Valdes-Mas R, Villamor N, Gutierrez-Abril J, Martin-Subero JI, Munar M, Rubio-Perez C, Jares P, Aymerich M, Baumann T et al.: Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* (2015) 526(7574):519–524. [PubMed: 26200345]
22. Klann TS, Black JB, Chellappan M, Safi A, Song L, Hilton IB, Crawford GE, Reddy TE, Gersbach CA: CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* (2017) 35(6):561–568. [PubMed: 28369033] ** This manuscript compared CRISPRi and CRISPRa screening across multiple cell lines, providing valuable information regarding different screening strategies.
23. Rajagopal N, Srinivasan S, Kooshesh K, Guo Y, Edwards MD, Banerjee B, Syed T, Emons BJ, Gifford DK, Sherwood RI: High-throughput mapping of regulatory DNA. *Nat Biotechnol* (2016) 34(2):167–174. [PubMed: 26807528] ** This study combines pooled screening with knock-in reporter assays to identify *cis*-regulatory elements.
24. Diao Y, Li B, Meng Z, Jung I, Lee AY, Dixon J, Maliskova L, Guan KL, Shen Y, Ren B: A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res* (2016) 26(3):397–405. [PubMed: 26813977]
25. Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O, Chen DD, Schupp PG, Vinjamur DS, Garcia SP, Luc S et al.: *BCL11A* enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* (2015) 527(7577):192–197. [PubMed: 26375006]
26. Diao Y, Fang R, Li B, Meng Z, Yu J, Qiu Y, Lin KC, Huang H, Liu T, Marina RJ, Jung I et al.: A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* (2017) 14(6):629–635. [PubMed: 28417999]
27. Simeonov DR, Gowen BG, Boontanart M, Roth TL, Gagnon JD, Mumbach MR, Satpathy AT, Lee Y, Bray NL, Chan AY, Lituiev DS et al.: Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* (2017) 549(7670):111–115. [PubMed: 28854172] ** This study used pooled CRISPRa screening to identify several CRISPRs-responsive elements (CaREs) with potential roles in autoimmunity.
28. Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, Pak RA et al.: A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* (2016) 167(7):1867–1882 e1821. [PubMed: 27984733]
29. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B et al.: Perturb-seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* (2016) 167(7):1853–1866 e1817. [PubMed: 27984732]
30. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I: Dissecting immune circuits by linking CRISPR- pooled screens with single-cell RNA-seq. *Cell* (2016) 167(7):1883–1896 e1815. [PubMed: 27984734]
31. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C: Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* (2017) 14(3):297–301. [PubMed: 28099430]
32. Xie S, Duan J, Li B, Zhou P, Hon GC: Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell* (2017) 66(2):285–299 e285. [PubMed: 28416141] ** This study pairs CRISPRi screens with scRNA-seq to interrogate enhancers for the first time.
33. Anderson EM, Haupt A, Schiel JA, Chou E, Machado HB, Strezoska Z, Lenger S, McClelland S, Birmingham A, Vermeulen A, Smith A: Systematic analysis of CRISPR- Cas9 mismatch tolerance reveals low levels of off-target activity. *J Biotechnol* (2015) 211(56–65. [PubMed: 26189696]
34. Hultquist JF, Schumann K, Woo JM, Manganaro L, McGregor MJ, Doudna J, Simon V, Krogan NJ, Marson A: A Cas9 ribonucleoprotein platform for functional genetic studies of HIV-host interactions in primary human T-cells. *Cell Rep* (2016) 17(5):1438–1452. [PubMed: 27783955]
35. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F: Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* (2014) 343(6166):84–87. [PubMed: 24336571]
36. Sanjana NE, Wright J, Zheng K, Shalem O, Fontanillas P, Joung J, Cheng C, Regev A, Zhang F: High-resolution interrogation of functional elements in the noncoding genome. *Science* (2016) 353(6307):1545–1549. [PubMed: 27708104]

37. Canver MC, Lessard S, Pinello L, Wu Y, Ilboudo Y, Stern EN, Needleman AJ, Galacteros F, Brugnara C, Kutlar A, McKenzie C et al.: Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait- associated loci. *Nat Genet* (2017) 49(4):625–634. [PubMed: 28218758] * This study demonstrates the use of multiple designer Cas9 nucleases and variant- specific library for interrogating noncoding functional elements.
38. Gasperini M, Findlay GM, McKenna A, Milbank JH, Lee C, Zhang MD, Cusanovich DA, Shendure J: CRISPR/Cas9-mediated scanning for regulatory elements required for *HPRT1* expression via thousands of large, programmed genomic deletions. *Am J Hum Genet* (2017) 101(2):192–205. [PubMed: 28712454]
39. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, Zwart W, Elkon R, Agami R: Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* (2016) 34(2):192–198. [PubMed: 26751173] ** This study used genomic catalogues and biochemical hallmarks to identify putative functional elements globally, and then conducted pooled CRISPR/Cas screens to identify transcription binding sites required for continued cancer proliferation.
40. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, Kane M, Cleary B, Lander ES, Engreitz JM: Systematic mapping of functional enhancer- promoter connections with CRISPR interference. *Science* (2016) 354(6313):769–773. [PubMed: 27708057] ** This is the first demonstration of CRISPRi screens for discovering regulatory elements and identifying their target genes.
41. Zhu S, Li W, Liu J, Chen CH, Liao Q, Xu P, Xu H, Xiao T, Cao Z, Peng J, Yuan P et al.: Genome-scale deletion screening of human long non-coding RNAs using a paired- guide RNA CRISPR-Cas9 library. *Nat Biotechnol* (2016) 34(12):1279–1286. [PubMed: 27798563] ** This study pioneered the use of paired sgRNA libraries for genome-wide deletion screens to study lncRNAs functions.
42. Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, Attenello FJ, Villalta JE, Cho MY, Chen Y, Mandegar MA et al.: CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* (2017) 355(6320). * This study introduced the use of CRISPRi-based screen to identify novel lncRNAs required for cellular growth.
43. Joung J, Engreitz JM, Konermann S, Abudayyeh OO, Verdine VK, Aguet F, Gootenberg JS, Sanjana NE, Wright JB, Fulco CP, Tseng YY et al.: Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* (2017) 548(7667):343–346. [PubMed: 28792927]
44. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG et al.: Genome-wide association study shows *BCL11A* associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A* (2008) 105(5):1620–1625. [PubMed: 18245381]
45. Sen DR, Kaminski J, Barnitz RA, Kurachi M, Gerdemann U, Yates KB, Tsao HW, Godec J, LaFleur MW, Brown FD, Tonnerre P et al.: The epigenetic landscape of T-cell exhaustion. *Science* (2016) 354(6316):1165–1169. [PubMed: 27789799]
46. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY: HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* (2016) 13(11):919–922. [PubMed: 27643841]
47. Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosh I, Sanjana NE, Shalem O et al.: A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* (2015) 162(3):675–686. [PubMed: 26189680]
48. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H et al.: Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* (2015) 160(6):1246–1260. [PubMed: 25748654]
49. Manguso RT, Pope HW, Zimmer MD, Brown FD, Yates KB, Miller BC, Collins NB, Bi K, LaFleur MW, Juneja VR, Weiss SA et al.: *In vivo* CRISPR screening identifies *PTPN2* as a cancer immunotherapy target. *Nature* (2017) 547(7664):413–418. [PubMed: 28723893]
50. Chow RD, Guzman CD, Wang G, Schmidt F, Youngblood MW, Ye L, Errami Y, Dong MB, Martinez MA, Zhang S, Renauer P et al.: AAV-mediated direct *in vivo* CRISPR screen identifies

functional suppressors in glioblastoma. *Nat Neurosci* (2017) 20(10): 1329–1341. [PubMed: 28805815]

51. Wallace J, Hu R, Mosbrugger TL, Dahlem TJ, Stephens WZ, Rao DS, Round JL, O'Connell RM: Genome-wide CRISPR-Cas9 screen identifies microRNAs that regulate myeloid leukemia cell growth. *PLoS One* (2016) 11 (4):e0153689. [PubMed: 27081855]
52. Gonzalez F, Zhu Z, Shi Z-D, Lelli K, Verma N, Li QV, Huangfu D: An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* (2014) 15(2):215–226. [PubMed: 24931489]
53. Mandegar MA, Huebsch N, Frolov EB, Shin E, Truong A, Olvera MP, Chan AH et al.: CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* (2016) 18(4):541–553. [PubMed: 26971820]

Figure 1. Pooled CRISPR screening strategies

In a pooled CRISPR screen, a lentiviral sgRNA library is transduced into a cell population at a low multiplicity of infection (MOI) with the goal that each cell receives only one sgRNA integration. Infected cells are put through a positive or negative selection process (i.e. drugs, virus infections, etc.) to identify sgRNAs that are either enriched or depleted, respectively, in the surviving cells through next-generation sequencing. One may also design the screen based on the effect of the sgRNA targeting on gene expression using FACS based on a fluorescent reporter or antibody staining: the relative sgRNA abundance in fluorescence-low (FL-low) versus fluorescence-high (FL-high) cell populations can be quantified. The sgRNAs and corresponding genes are then ranked based on the degree of sgRNA depletion or enrichment, and potential candidate gene hits can then be further validated.

Table 1:

Summary of CRISPR screens targeting the noncoding genome:

Publication (year)	Perturbation Methods	sgRNA Libraries	Target Areas	Cell Types	Assay Methods	References
Canver et al. (2015)	Cas9 nuclease	702 sgRNAs targeting 3 DHSs and non-targeting controls	Total of 4-kb sequences encompassing putative <i>BCL11A</i> enhancers	HUDEP-2 human umbilical cord blood-derived erythroid progenitor cell line	FACS-based on antibody staining	[25]
Korkmaz et al. (2016)	Cas9 nuclease	1,116 sgRNAs for <i>p53</i> enhancer and 97 sgRNAs for <i>ERα</i> enhancer	685 <i>p53</i> and 73 <i>ERα</i> bound enhancer regions	Human BJ fibroblast cells, MCF-7 and T47D breast cancer cell line	Cell proliferation	[39]
Sanjana et al. (2016)	Cas9 nuclease	18,315 sgRNAs (6,682 sgRNAs for <i>NF1</i> , 6,934 sgRNAs for <i>NF2</i> and 4,699 sgRNAs for <i>CUL3</i>)	100-kb upstream and downstream of <i>NF1</i> , <i>NF2</i> and <i>CUL3</i> each	A375 human melanoma cell line	Cell survival after vemurafenib treatment	[36]
Rajagopal et al. (2016)	Cas9 nuclease	3,908 tiling sgRNAs for each of <i>Tdgtf1</i> , <i>Nanog</i> , <i>Rpp25</i> , and <i>Zip42</i>	Upstream regulatory regions of <i>Tdgtf1</i> (40-kb upstream), <i>Nanog</i> , (20-kb upstream + 92-Mb distal region) <i>Rpp25</i> and <i>Zip42</i> (upstream regulators based on ChIA-PET)	Mouse embryonic stem cells	FACS-based on knock-in reporter	[23]
Diao et al. (2016)	Cas9 nuclease	1,964 sgRNAs based on DHSs and ChIP-seq	174 candidate regulatory regions within 1-Mb of <i>POU5F1</i>	H1 human embryonic stem cells	FACS-based on transgenic reporter	[24]
Fulco et al. (2016)	dCas9-KRAB	98,000 tiling sgRNAs	Total of ~1.29-Mb sequences around <i>GATA1</i> and <i>MYC</i>	K562 human erythroleukemia cell line	Cell proliferation	[40]
Zhu et al. (2016)	Cas9 nuclease and dCas9-KRAB/dCas9-VP64 for validation	12,472 sgRNA pairs	647 lncRNA loci	Huh7.5 hepatocarcinoma cell line	Cell proliferation	[41]
Wallace et al. (2016)	Cas9 nuclease	GeCKO v2 library (123,411 sgRNAs for 19,050 protein coding genes), which also include 7,456 sgRNAs targeting 1,864 miRNAs	1,864 miRNAs (4 sgRNA/miRNA) loci	MV4-11 myeloid leukemia cell line	Cell proliferation	[51]
Diao et al. (2017)	Cas9 nuclease	11,570 sgRNA pairs	2-Mb of <i>POU5F1</i> cis-regulatory region	H1 human embryonic stem cells	FACS-based on transgenic reporter	[26]
Xie et al. (2017)	dCas9-KRAB	51,448 sgRNAs	71 targeted regions across 15 super enhancers	K562 human erythroleukemia cell line	scRNA-seq	[32]
Simeonov et al. (2017)	dCas9-VP64	10,780 sgRNAs for <i>CD69</i> and 20,412 sgRNAs for <i>IL2RA</i>	135-kb across <i>CD69</i> and total of 178-kb upstream and downstream of <i>IL2RA</i> locus	Jurkat human T cell line	FACS-based on antibody staining	[27]
Klann et al. (2017)	dCas9-KRAB and dCas9-p300	12,472 sgRNAs targeting β -globin DHSs and 12,422 sgRNAs for <i>HER2</i> DHSs	4.5-Mb and 4-Mb, respectively, surrounding the β -globin and <i>HER2</i> loci	K562 human erythroleukemia cell line, HEK293T, A431 human	FACS-based on knock-in reporter	[22]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Publication (year)	Perturbation Methods	sgRNA Libraries	Target Areas	Cell Types	Assay Methods	References
				epidermoid carcinoma cell line		
Liu et al. (2017)	dCas9-KRAB	164,010 sgRNAs	16,401 lncRNA loci	Cancer cell lines (K562, HeLa, U87, MCF7, MDA-MB-231) and human induced pluripotent stem cells	Cell proliferation	[42]
Joung et al. (2017)	dCas9-SAM	~96,000 sgRNAs	10,504 lncRNA loci	A375 human melanoma cell line	Cell survival after vemurafenib treatment	[43]
Chow et al. (2017)	Cas9 nuclease	288 sgRNAs	5 sgRNA each against 56 genes (49 tumor suppressor and 7 housekeeping genes) plus 8 nontargeting sgRNAs	<i>In vivo</i> screen using LSL-Cas9-CFP mice (for conditional Cas9 expression)	Cell proliferation	[50]