Trends in Genetics

CellPress

Forum

CRISPR Screens to Discover Functional Noncoding Elements

Jason B. Wright[1,2] and Neville E. Sanjana[3,4,*]

A major challenge in genomics is to identify functional elements in the noncoding genome. Recently, pooled clustered regularly interspersed palindromic repeat (CRISPR) mutagenesis screens of noncoding regions have emerged as a novel method for finding elements that impact gene expression and phenotype/disease-relevant biological processes. Here we review and compare different approaches for high-throughput dissection of noncoding elements.

Which Sequences in the Genome Impact Human Biology?

Less than 2% of the ~3 billion DNA base pairs in the human genome encode proteins, whereas most of the human genome comprises noncoding regions. The function of noncoding regions is less well understood than the coding genome and, for many noncoding regions, there is vigorous debate about whether they have any function at all [1–3]. Nonetheless, sequence conservation estimates have found that ~10% of the genome is under selection, supporting a functional role for some noncoding sequences [4]. Identifying functional elements in the vast noncoding space and understanding their roles in different biological process is one of the major current challenges in genomics.

Inroads into this challenge have been made over the past two decades through the characterization of biochemical hallmarks that correlate with putative noncoding functional elements, such as chromatin accessibility, chromatin conformation, transcription factor binding prediction, epigenetic modifications, and conservation. Recent consortium efforts like the Encyclopedia of DNA Elements (ENCODE) and the Roadmap Epigenomics Project have produced vast quantities of genome-scale data that are widely used to predict regulatory function [1,5]. Diverse types of gene regulatory elements such as promoters, enhancers, and functional noncoding RNAs hint at the presence of a complex noncoding landscape but presently these features provide only hypotheses about function, not proof of a role in biological processes [6]. Large-scale assays for noncoding function, such as massively parallel reporter assays (MPRAs), place small, synthesized 100–200-bp putative functional elements before a minimal promoter and quantify mRNA expression [7]. MPRAs have recently been employed to quantify and compare thousands of expression-modulating variants [8,9] but have several limitations. Since the assay uses episomal reporters, analyzed variants lack native chromatin context and other surrounding genome features. Also, due to the mRNA readout, it is not feasible to detect variants that work via post-transcriptional or feedback mechanisms.

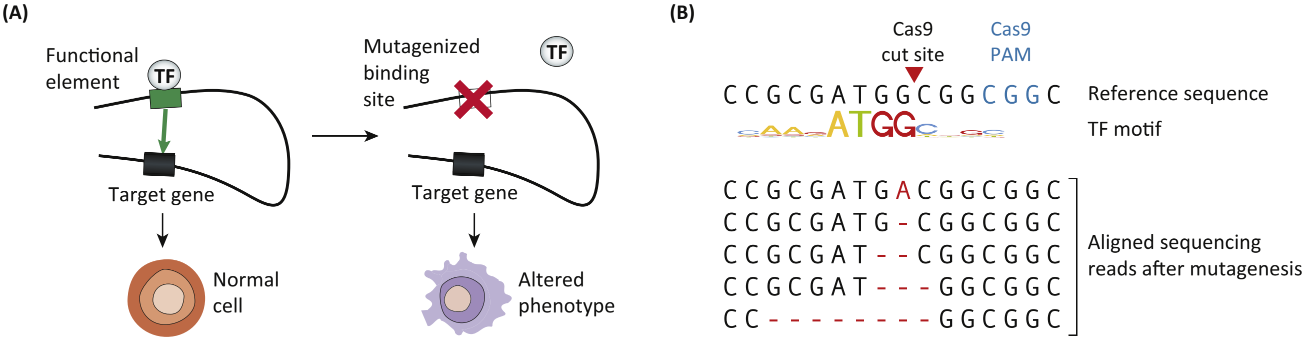
A more direct approach for identifying functional elements is to modify or mutagenize an element in its native context and see whether changes in gene expression or cellular function follow (Figure 1A). Until recently, genome editing has been challenging in human cells. Over the past few years, RNA-guided nucleases derived from CRISPR microbial immune systems (e.g., Cas9 from *Streptococcus pyogenes*) have enabled high-throughput genome modification in cells and tissues from diverse organisms [10]. CRISPR systems are targeted to different genomic sequences by a short single guide RNA (sgRNA), enabling rapid synthesis of large libraries of CRISPR reagents using array oligonucleotide synthesizers similar to those used for genotyping arrays. For protein-coding genes, loss-of-function and gain-of-function screens using genome-scale CRISPR libraries with thousands of sgRNAs have identified genes involved in diverse cellular phenotypes, including cell survival, drug/toxin resistance, immune signaling, and cancer metastasis [11,12].

Recently, genome engineering techniques have been applied to the noncoding genome. Targeted mutations in noncoding regions can result in disruption of functional elements such as promoter or enhancer sites (Figure 1B). Many of these regions are inaccessible to manipulation by other pooled screening techniques like RNAi. However, there are challenges even when working at the DNA level. For example, in coding regions any frameshift mutation can result in loss of function but disruption of smaller noncoding elements might require mutagenesis at a precise location (e.g., a 5–10-bp transcription factor-binding site).

Although there are differences in library design and phenotypic selection between different CRISPR screens, all employ libraries of sgRNAs to identify functional elements within noncoding regions (Figure 2). Here we review and compare several recent noncoding CRISPR screens and examine how genome engineering can further our understanding of the noncoding genome.

Targeted Screens Guided by Disease Genetics

Genome-wide genetic association studies (GWASs) have revealed thousands of variants that correlate with human disease and the vast majority lie in noncoding regions, implying that regulatory variation is an important component of inherited disease risk. However, finding the exact causal variant among other variants can be challenging due to linkage. An early example of a noncoding screen identifying a causal variant is in hemoglobin regulation. An intronic variant in the gene *BCL11A* was identified by a GWAS as an ameliorating factor in β-thalassemia and sickle-cell anemia [13]. These disorders are commonly due to defects in the


(A)
Functional element
TF
Target gene
Normal cell
Mutagenized binding site
TF
Target gene
Altered phenotype
(B)
Cas9 cut site
Cas9 PAM
C C G C G A T G G C G G C G G C
Reference sequence
ATGG
TF motif
C C G C G A T G A C G G C G G C
C C G C G A T G - C G G C G G C
C C G C G A T - - C G G C G G C
C C G C G A T - - - G G C G G C
C C - - - - - - - - - G G C G G C
Aligned sequencing reads after mutagenesis

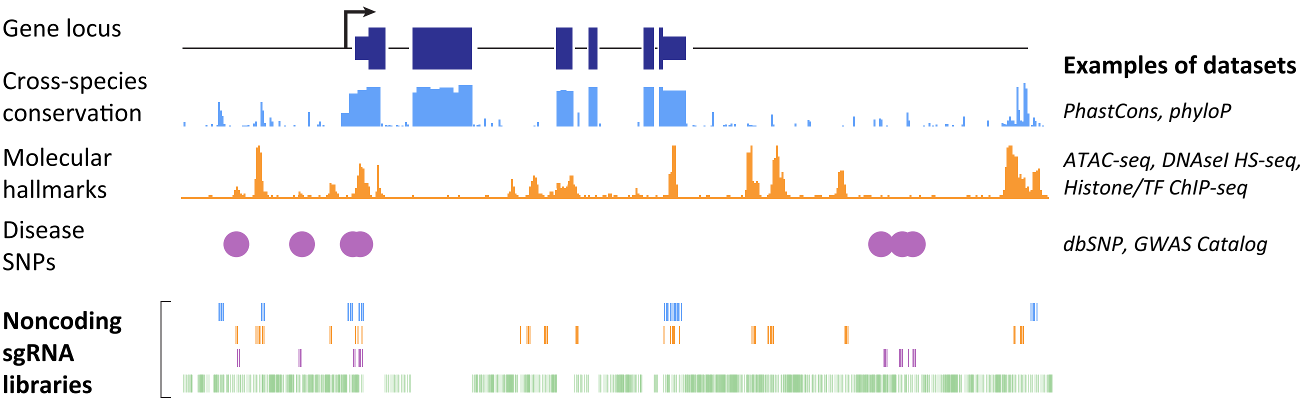
Trends in Genetics

Figure 1. Targeted Clustered Regularly Interspersed Palindromic Repeat (CRISPR) Mutagenesis Disrupts Noncoding Functional Elements via Insertion–Deletion (Indel) Mutations. (A) Mutagenesis of a transcription factor (TF)-binding site that lies distal to a promoter of a gene. After targeted mutagenesis at the distal site, the TF no longer recognizes the sequence and no longer binds, resulting in altered gene expression and cellular phenotype. (B) Mutagenesis of a canonical TF motif (in this example, for YY1) with a single guide RNA (sgRNA). The sequenced alleles from the resulting polygenic population reflect the diversity in nonhomologous end-joining double-strand break repair outcomes after Cas9 nuclease activity. None of the post-genome modification alleles matches the maximum-likelihood YY1-binding motif. (B) adapted from [18].

adult form of hemoglobin and it has been shown that loss of *BCL11A* results in derepression of fetal hemoglobin (which substitutes for the adult form). However, within this large intron of *BCL11A*, the precise location of the functional element that modulates *BLC11A* expression was unknown.

Mutagenesis of all Cas9-targetable sites in three erythroid-specific enhancer regions (identified previously by a biochemical hallmark of noncoding function – DNase I hypersensitivity) in the *BCL11A* intron was performed using a pooled CRISPR library to find the causal variant responsible for controlling *BCL11A* expression [14]. A total of 702 sgRNAs were screened in HUDEP-2 cells, which usually express only low levels of fetal hemoglobin. After genome modification, cells were labeled with an antibody to fetal hemoglobin and sorted using fluorescence-assisted cell sorting to enrich for cells with a high level of fetal hemoglobin. The top-scoring sgRNAs mapped to a *GATA1*-binding site; GATA1, a master regulator of erythropoiesis, acts as an enhancer of *BCL11A* in *cis*. Thus, mutations (or natural variants) at this site reduce *BCL11A* expression.

Since complete loss of *BCL11A* is lethal, it is a challenging drug target for treating hemoglobinopathies. However, given that the intronic, GATA1-binding enhancer identified in the screen acts in a cell type-specific manner, there is great potential for therapeutic genome editing



Trends in Genetics

Figure 2. Noncoding Clustered Regularly Interspersed Palindromic Repeat (CRISPR) Libraries Can Be Designed with Single Guide RNAs (sgRNAs) that Target Known Genome Features or for Saturation Mutagenesis. In a noncoding region, different genomic features associated with the region (top) and four different noncoding CRISPR libraries (bottom) are depicted. Blue sgRNAs target sites of high cross-species conservation. Orange sgRNAs target sites with a specific molecular hallmark (e.g., transcription factor-binding sites, histone modifications). Purple sgRNAs target known human SNPs with disease or phenotype associations. Green sgRNAs target as many genomic locations as possible in an unbiased fashion over the noncoding region.

at this noncoding locus to modulate *BCL11A* in erythroid cells alone.

Targeted Screens Guided by Biochemical Hallmarks

Another approach for selecting putative functional elements is to use genomic catalogs of biochemical hallmarks, such as ENCODE consortium data. By targeting hallmarks such as transcription factor binding sites or peaks of post-translational histone modifications, it is possible to nominate putative functional elements. Using this strategy, two CRISPR libraries were designed based on prior ChIP-seq datasets of *p53* and estrogen receptor alpha (*ERα*) – transcription factors commonly mutated in cancers – and these libraries were used to identify binding sites required for continued cancer proliferation (Figure 1B) [15].

By targeting ~700 *p53*-binding sites using 1116 sgRNAs in an inducible mutant-RAS model of oncogene-induced senescence, they identified key sites required for p53-based senescence after mutant RAS induction. Two sites near *CDKN1A*, a well-known mediator of p53-induced cell cycle arrest, were enriched in the initial screen (enrichment for continued proliferation and escape from senescence after CRISPR mutagenesis at these sites) and further validated using cell cycle assays and a second, higher-resolution CRISPR screen over a smaller region. The authors also performed a proliferation screen in ER-dependent breast cancer cell lines using a small library of 97 sgRNAs targeting ERα-binding sites. Since cell growth depends on ERα binding, disruption of critical binding sites prevents further proliferation. For this screen, the ERα sites targeted by depleted sgRNAs were scored as pro-growth elements. Previous chromatin-interaction analysis by paired-end tag (ChIA-PET) showed that one ERα enhancer identified from the screen physically interacts with the G1 cell cycle gene *CCND1*, which is commonly overexpressed in breast cancer.

Although the authors focused on the role of two specific types of transcription factor binding site in cancer evolution, this approach can be used for any biochemical hallmark genome wide. For example, genome-wide CRISPR libraries targeting miRNAs have been used to uncover miRNAs that drive cancer metastasis *in vivo* [16]. In this manner, noncoding elements that exert phenotypic effects through multiple genes (e.g., all gene targets of a miRNA) can also be investigated in a high-throughput fashion.

Screens for Control of Gene Expression

A major role of noncoding genome elements is the control of gene expression, which can itself be used as a selectable phenotype for pooled screens. To perform a noncoding screen for gene expression, GFP knock-in mouse embryonic stem cell lines were generated for four different genes: *Nanog*, *Rpp25*, *Tdgf1*, and *Zfp42* [17]. For each reporter line, the authors delivered sgRNAs targeting regulatory regions near each gene using a homologous recombination-based approach. They identified several *cis*-regulatory elements that changed GFP expression when mutated but that do not coincide with the typical predictive biochemical features of enhancer elements such as H3K4me1 and H3K27ac. By sequencing mutations after genome modification at these locations, they mapped the key sequence features of the novel regulatory elements. An attractive feature of this screening paradigm is that it can be expanded to any expressed gene for which an appropriate fluorescent reporter line is available.

Unbiased Screens for Functional Elements that Impact Disease

Ideally, functional screens should interrogate large genomic intervals in an unbiased fashion while taking advantage of disease-relevant phenotypes for selective enrichment of relevant sgRNAs from the pool. A recent study looked for functional elements in regions surrounding genes that mediate resistance to the BRAF inhibitor vemurafenib in a BRAF-mutant melanoma cell line [18]. These genes were identified in a previous genome-wide (coding) CRISPR screen using the same drug-resistance phenotype. For the noncoding screens, ~18 000 sgRNAs were targeted densely across 200–300 kb of sequence flanking each of the genes. In these libraries, the average distance between neighboring sgRNA target sites was ~15 bp, which was about the same length as the average indel mutation, enabling high-density coverage of the noncoding regions.

After vemurafenib selection, enriched target sites overlap with melanoma-specific open chromatin regions (from DNase I HS-seq and ATAC-seq profiles) and with regions of evolutionary conservation. Chromatin conformation capture (3C) analysis indicated that noncoding regions that tend to physically interact with the gene promoter via chromatin loops show strong enrichment for functional regulatory elements. For many of the enriched regions, mutations at these sites cause a significant decrease in general biochemical hallmarks of regulatory activity, such as H3K27ac for distal enhancers and H3K4me3 for promoter-proximal regions. At specific sgRNA target sites where bioinformatically identified binding motifs exist, several transcriptional factors previously implicated in melanoma, such as JUN, FOS, YY1, ZNF263, and CTCF, had decreased binding after CRISPR mutagenesis (Figure 1B).

This study demonstrates the effectiveness of pairing a genome-wide screen over coding regions with subsequent high-resolution, unbiased noncoding screens around genes of interest identified in the coding screen. Mutagenesis of several of the regulatory elements from the noncoding screen results in a phenotypic impact of similar magnitude to that of coding mutations, emphasizing the importance of noncoding variants to disease biology.

Prospects of Noncoding Functional Screens

Rapid advances and innovative applications of high-throughput mutagenesis screens have opened new avenues for both hypothesis-driven and unbiased interrogation of noncoding sequences. Presently, noncoding screens have been confined to mutagenesis over regions of 10 kb to 1 Mb. Scaling up to screen over larger genomic intervals (or even entire genomes) will require new tools for creating large deletions or programmed rearrangements (e.g., targetable recombinases or transposases). Continued advances in genome engineering tool development and high-throughput phenotypic screens have great potential for exploring the basic functional architecture of the noncoding genome.

Acknowledgements

The authors thank R. Macrae, R. Gupta, and J. Alfoldi for critical reading of the manuscript and F. Zhang for support and advice. J.B.W. is supported by the NIH through a Ruth L. Kirschstein National Research Service Award (F32-DK096822). N.E.S. is supported by the NIH through a NHGRI Pathway to Independence Award (R00-HG008171).

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
[2]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]New York Genome Center, New York, NY 10013, USA
[4]Center for Genomics and Systems Biology, Department of Biology, New York University, NY 10003, USA

*Correspondence: nsanjana@nygenome.org (N.E. Sanjana).

http://dx.doi.org/10.1016/j.tig.2016.06.004

References

1. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
2. Graur, D. *et al.* (2013) On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5, 578–590
3. Eddy, S.R. (2013) The ENCODE project: missteps overshadowing a success. *Curr. Biol.* 23, R259–R261
4. Rands, C.M. *et al.* (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
5. Roadmap Epigenomics Consortium *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330
6. Kellis, M. *et al.* (2014) Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6131–6138
7. Melnikov, A. *et al.* (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30, 271–277
8. Tewhey, R. *et al.* (2016) Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* 165, 1519–1529
9. Ulirsch, J.C. *et al.* (2016) Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* 165, 1530–1545
10. Hsu, P.D. *et al.* (2014) Development and applications of CRISPR–Cas9 for genome engineering. *Cell* 157, 1262–1278
11. Shalem, O. *et al.* (2015) High-throughput functional genomics using CRISPR–Cas9. *Nat. Rev. Genet.* 16, 299–311
12. Sanjana, N.E. (2016) Genome-scale clustered regularly interspaced short palindromic repeats pooled screens. *Anal. Biochem.* Published online June 1, 2016. http://dx.doi.org/10.1016/j.ab.2016.05.014
13. Uda, M. *et al.* (2008) Genome-wide association study shows *BCL11A* associated with persistent fetal hemoglobin and amelioration of the phenotype of β-thalassemia. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1620–1625
14. Canver, M.C. *et al.* (2015) *BCL11A* enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature* 527, 192–197
15. Korkmaz, G. *et al.* (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR–Cas9. *Nat. Biotechnol.* 34, 192–198
16. Chen, S. *et al.* (2015) Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246–1260
17. Rajagopal, N. *et al.* (2016) High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* 34, 167–174
18. Sanjana, N.E. *et al.* (2016) High-resolution interrogation of functional elements in the noncoding genome. *BioRxiv.* Published online April 18, 2016. http://dx.doi.org/10.1101/049130