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

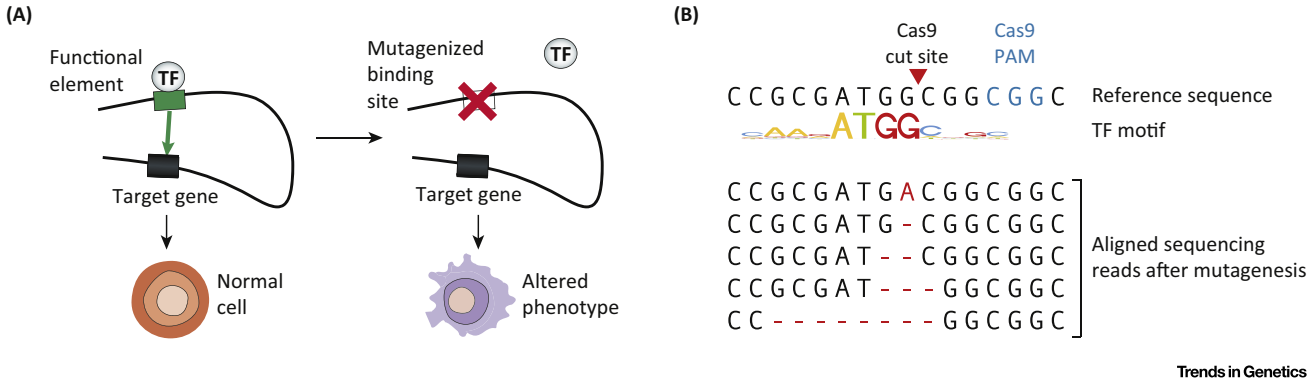

Figure 1. Targeted Clustered Regularly Interspersed Palindromic Repeat (CRISPR) Mutagenesis Disrupts Noncoding Functional Elements via Insertion–Deletion (Indel) Mutations. (A) Mutagenesis of a transcription factor (TF)-binding site that lies distal to a promoter of a gene. After targeted mutagenesis 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

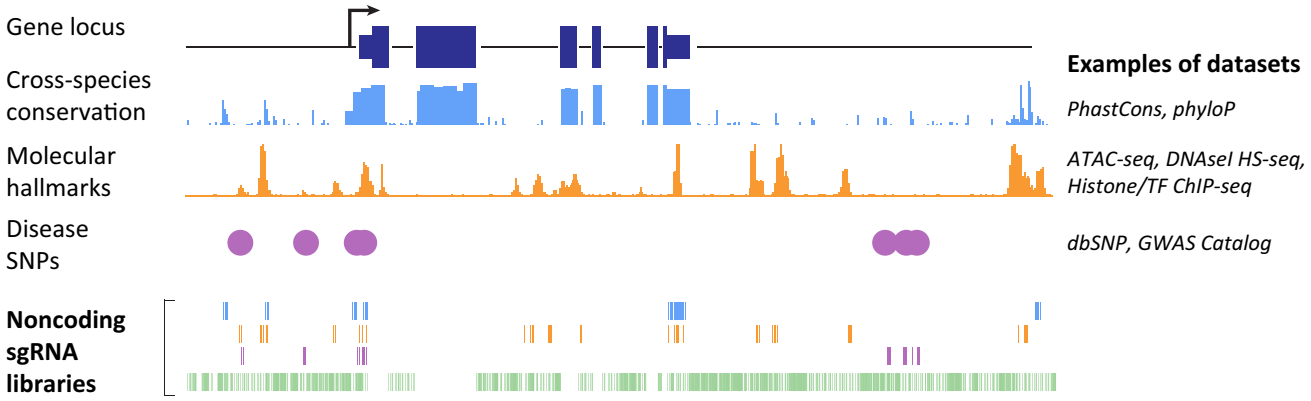

Figure 2. Noncoding Clustered Regularly Interspersed Palindromic Repeat (CRISPR) Libraries Can Be Designed with Single Guide RNAs (sgRNAs) that Target Known Genome Features or for Saturation Mutagenesis. In a noncoding region, different genomic features associated with the region (top) and four different 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.

Download English Version:

https://daneshyari.com/en/article/2824580

Download Persian Version:

https://daneshyari.com/article/2824580

Daneshyari.com