Mapping regulatory elements

Yuexin Zhou & Wensheng Wei

Regulatory genomic elements can now be studied in their native context using two CRISPR-based high-throughput approaches.

Regulatory regions of the genome have long taken a back seat to protein-coding regions in the search for disease-causing mutations. But evidence that they contribute to a wide array of diseases has continued to mount, together with a need for improved methods to characterize their function. In this issue, two groups describe CRISPR-Cas9 genetic screens for interrogating the native function of enhancers[1] and *cis*-regulatory elements[2] in a high-throughput manner. The studies provide interesting insights into the roles of the noncoding genome and suggest that CRISPR-Cas9 offers considerable advantages over competing tools to investigate regulatory elements.

Recent reports describing the effects of promoters and enhancers in tumorigenesis and other disease mechanisms[3,4] have intensified interest in genomic regulatory elements, but the staggeringly large number of such elements in the mammalian genome still defies analysis. The first high-throughput methods based on massively parallel reporter assays relied on cloning regulatory elements into plasmid vectors[5–7]. Although a much-needed step in the right direction, these approaches could not capture the function of elements in their native genomic context.

CRISPR-Cas9 systems have been used to study gene function on a large scale[8–11] and, in a recent report[12], to dissect enhancer function of a single gene through CRISPR-Cas9–mediated saturation mutagenesis. The papers by Korkmaz *et al.*[1] and Rajagopal *et al.*[2] in this issue take the next step by enabling high-throughput mapping of regulatory elements, including enhancers[1,2] and unmarked regulatory elements[2], at single-base resolution in their native context (**Fig. 1**).

Korkmaz *et al.*[1] developed two CRISPR-Cas9 genetic screens to identify enhancer elements and their target genes. They focus specifically on enhancers bound by p53 and estrogen receptor alpha (ERα), and search for these predicted regions in ENCODE, ChIP-seq and GRO-seq data sets. They design 1,116 and 97 sgRNAs to target the predicted enhancer regions of p53 and ERα, respectively, and then construct cell libraries through lentiviral delivery of Cas9 and sgRNAs. Using this approach, they identify several candidate p53- and ERα-bound enhancers by either positive or negative CRISPR-Cas9 library screens. p53-dependent oncogene-induced senescence in BJ-RAS^{G12V} cells is used to search for potential p53-bound enhancers through a positive screen, whereas ERα-dependent cell proliferation is used to identify ERα-bound enhancers in three distinct cell lines in a dropout (negative) screen. After culturing the cells for several weeks, the authors sequence candidate sgRNAs that were either positively (for p53) or negatively (for ERα) enriched in the experimental library

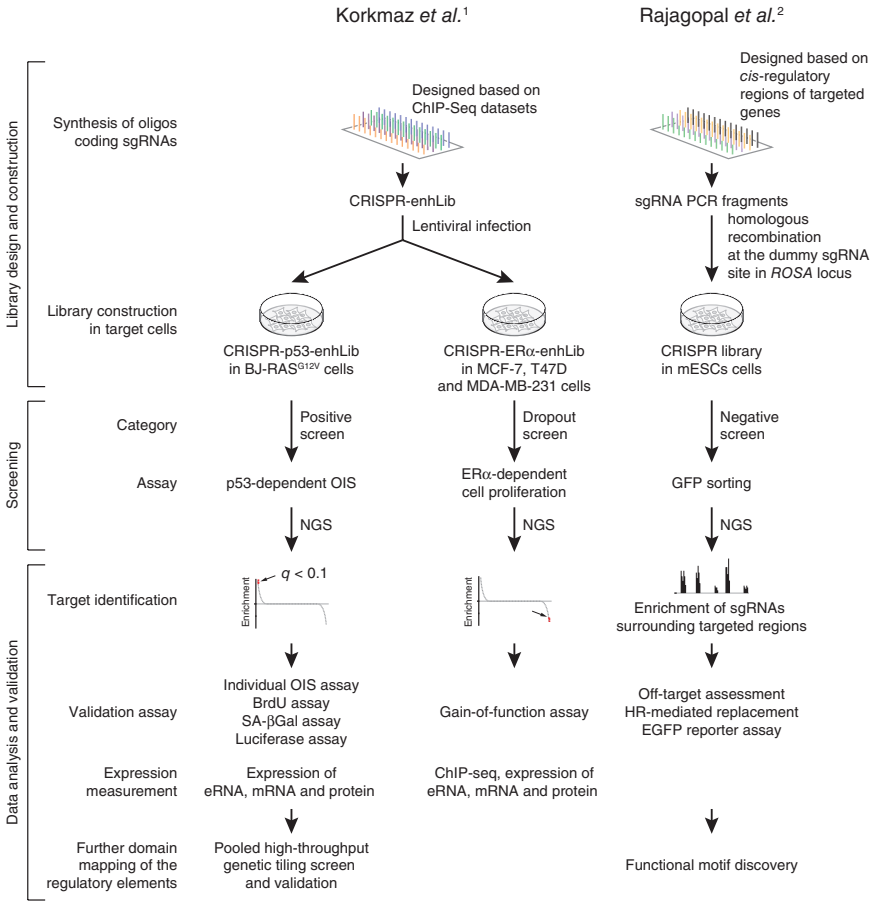

Figure 1 Comparison of two approaches for functional screening of regulatory elements of target genes. The common steps include library design and construction, screening, data analysis and validation. Korkmaz *et al.*[1] design sgRNAs based on ChIP-seq data sets, generate the cell libraries via lentiviral infection, and identify p53- and ERα-bound enhancers through either positive or negative screens using p53-dependent oncogene-induced senescence and ERα-dependent cell proliferation assays, respectively. Rajagopal *et al.*[2] tile sgRNAs in *cis*-regulatory regions surrounding target genes, construct cell libraries through homologous recombination–mediated knock-in of sgRNAs at the dummy sgRNA site in the *ROSA* locus, and screen the library using a GFP reporter. Both papers conduct a series of validation experiments to confirm screening results and map functional domains of the regulatory elements at high resolution. OIS, oncogene-induced senescence; BrdU, bromodeoxyuridine; SA-βGAL, senescence associated β-galactosidase; EGFP, enhanced green fluorescent protein; eRNA, enhancer-associated RNA; HR, homologous recombination; NGS, next-generation sequencing.

Yuexin Zhou and Wensheng Wei are at the Biodynamic Optical Imaging Center (BIOPIC), Peking-Tsinghua Center for Life Sciences, the State Key Laboratory of Protein and Plant Gene Research, Beijing, China, and the School of Life Sciences, Peking University, Beijing, China.
e-mail: wswei@pku.edu.cn

cells. A series of validation assays confirms that the newly identified p53- and ERα-bound enhancers both have important roles in regulating their corresponding genes (**Fig. 1**, left).

In a different approach, Rajagopal *et al.*[2] designed a screen named multiplexed editing regulatory assay (MERA), which, unlike traditional lentiviral-based CRISPR-Cas library construction, uses a unique cloning-free strategy to ensure that a single gRNA is incorporated into cells without the need for plasmid construction (**Fig. 1**). They achieve this by integrating a dummy gRNA into prepositioned sites in the genome and then replacing the dummy gRNA with a pooled library of gRNAs through CRISPR-Cas9–based homologous recombination. In this way, they are able to tile gRNAs and induce mutations across the *cis*-regulatory region of a GFP-tagged locus. They can then directly measure the effect of mutated regulatory regions on gene expression levels.

Rajagopal *et al.*[2] use MERA to investigate the contribution of *cis*-regulatory elements to the expression of four genes specific to mouse embryonic stem cells: *Nanog*, *Rpp25*, *Tdgf1* and *Zfp42*. To identify the gRNAs that resulted in the loss of GFP expression, they enrich populations of cells with low or no target gene expression after prolonged cell culturing using flow cytometry. By sequencing gRNAs from these cell populations, they uncover predicted regulatory elements in all four genes, including promoters, enhancers and transcription factor (TF) binding sites. Interestingly, they identify a new class of regulatory elements, which they designate unmarked regulatory elements (UREs), that are shown to affect gene expression but do not contain known markers of regulatory activity. They also validate their results, analyze off-target effects, and further map functional motifs within regulatory elements by examining base variation from sequencing data (**Fig. 1**).

These two studies establish powerful approaches to systematically explore the function of regulatory elements. Although they differ in the screening and validation assays used and in the genes targeted, there are a number of similarities. These similarities include the knowledge-based prediction of regulatory regions, sgRNA design and synthesis, library construction and screening, and the use of sequence analysis and candidate validation (**Fig. 1**). Both groups produce convincing results, uncover functional motifs in regulatory elements and identify novel enhancer elements or even previously unmarked regulatory elements.

The studies also present some unique methodological characteristics. First, Korkmaz *et al.*[1] design sgRNAs based on existing ChIP-seq data sets and focus only on enhancer elements, whereas Rajagopal *et al.*[2] tile sgRNAs across a broad range of *cis*-regulatory regions surrounding target genes. Second, Korkmaz *et al.* deliver sgRNAs into cells by conventional lentivirus-based infection, whereas Rajagopal *et al.* first construct a dummy sgRNA integrated cell line, followed by its replacement with library sgRNAs through homologous recombination. Although the latter library-construction method avoids the need for plasmid construction and has better stringency to ensure that only one sgRNA is integrated into each cell, its efficiency depends heavily on homologous recombination activity, which limits cell choices. Finally, whereas Rajagopal *et al.*[2] rely on the pre-generated GFP knock-in as a reporter for library screening, Korkmaz *et al.*[1] take advantage of phenotypic change caused directly by target-gene expression, which represents a natural way to monitor native gene expression.

Both reports[1,2] showcase a high level of efficiency and robustness in the study of genetic regulatory elements in their native context. The physiological coverage sizes and the complexity of sequences are all hugely advantageous over traditional methods. Despite these advantages, the methods do have several limitations. sgRNAs are designed to target known or predetermined genomic regions, which reduces the chances of uncovering unknown regulatory elements. It is worth noting that some *cis* elements might work redundantly and that the small size of indels at a single locus may not give rise to any phenotypic change. Modifying the library design to use paired gRNAs to generate mutations at two loci or to create large-fragment deletions could potentially resolve this problem. Furthermore, the choice of sgRNAs depends heavily on the frequency of protospacer-adjacent motifs (PAMs), and the coverage density in certain areas might be too scarce for a saturation screen. It may therefore be necessary to combine different types of CRISPR systems to improve the screen resolution. Moreover, the approaches create a unique library for each individual screen, and it would be preferable to construct a universal library for unbiased screening.

A final consideration concerns throughput. Both studies focus on one or a few target genes, and the sizes of the libraries are still relatively small. The number of predicted enhancers in the human genome is over half a million[1], without counting the unmarked elements. If we are to have any hope of studying these elements and their combinatorial actions comprehensively, further technological improvements in the coverage and speed of screening are needed.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Korkmaz, G. *et al. Nat. Biotechnol.* **34**, 192–198 (2016).
2. Rajagopal, N. *et al. Nat. Biotechnol.* **34**, 167–174 (2016).
3. Mansour, M.R. *et al. Science* **346**, 1373–1377 (2014).
4. Vahedi, G. *et al. Nature* **520**, 558–562 (2015).
5. Patwardhan, R.P. *et al. Nat. Biotechnol.* **30**, 265–270 (2012).
6. Melnikov, A. *et al. Nat. Biotechnol.* **30**, 271–277 (2012).
7. Sharon, E. *et al. Nat. Biotechnol.* **30**, 521–530 (2012).
8. Shalem, O. *et al. Science* **343**, 84–87 (2014).
9. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. *Science* **343**, 80–84 (2014).
10. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.C. & Yusa, K. *Nat. Biotechnol.* **32**, 267–273 (2014).
11. Zhou, Y. *et al. Nature* **509**, 487–491 (2014).
12. Canver, M.C. *et al. Nature* **527**, 192–197 (2015).

Personalized nutrition through big data

Daniel McDonald, Gustavo Glusman & Nathan D Price

A systems model of glycemic response identifies individually tailored diets that keep blood sugar in check.

Digital phenotypes generated through omics assays and wearable devices are poised to change the face of healthcare, but so far evidence that they can provide predictive dietary recommendations for individuals has been lacking. Even anticipating the effect of meals on blood glucose levels is difficult, and this is usually estimated using glycemic indices, which are imprecise[1] and generic. And although the composition of the human gut microbiota has been linked to diabetes, cause

Daniel McDonald, Gustavo Glusman & Nathan D. Price are at the Institute for Systems Biology, Seattle, Washington, USA.
e-mail: nprice@systemsbiology.org