Published in final edited form as:

Adv Exp Med Biol. 2019 ; 1185: 359–364. doi:10.1007/978-3-030-27378-1_59.

High-Throughput Analysis of Retinal Cis-Regulatory Networks by Massively Parallel Reporter Assays

Inez Y. Oh,

Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, Saint Louis, MO, USA

Shiming Chen

Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, Saint Louis, MO, USA

Department of Developmental Biology, Washington University School of Medicine, Saint Louis, MO, USA

Abstract

Inherited retinal degenerations are diverse and debilitating blinding diseases. Genetic tests and exome sequencing have identified mutations in many protein-coding genes associated with such diseases, but causal sequence variants remain to be found in many retinopathy cases. Since 99% of our genome does not code for protein but contains *cis*-regulatory elements (CREs) that regulate the expression of essential genes, CRE variants might hold the answer for some of these cases. However, identifying functional CREs within the noncoding genome and predicting the pathogenicity of CRE variants pose a significant challenge. Here, we review the development of massively parallel reporter assays in the mouse retina, its use in dissecting retinal *cis*-regulatory networks, and its potential application for developing therapies.

Keywords

Noncoding DNA; *Cis*-regulatory element; High-throughput functional analysis; Transcriptional regulation; Massively parallel reporter assay (MPRA); CRE-seq; Retina transcription factor; CRX; NRL

59.1 Introduction

Inherited retinal degenerations (IRDs) are debilitating, often blinding, conditions that affect approximately 100,000 people in the USA. Over 200 causal genes have been discovered (RetNet), but cannot account for all IRD cases. Specific noncoding genomic sequences, known as *cis*-regulatory elements (CREs), mediate transcription factor (TF) action to coordinate coding gene expression (Shlyueva et al. 2014). Since many IRD cases remain unexplained despite diagnostic exome sequencing (Siemiatkowska et al. 2014), CRE variants might hold the answer for these cases.

Identifying CREs and predicting the pathogenicity of CRE variants pose a major challenge. Early studies used interspecies sequence conservation as a proxy for functional conservation (Visel et al. 2007) to identify putative CREs in 5' and 3' flanking sequences and introns of their target genes (Kleinjan and van Heyningen 2005). These putative CREs were then individually cloned into reporter vectors to test their ability to drive gene expression in cell-based luciferase assays or in the expected spatiotemporal pattern in transgenic mice. However, such one-by-one reporter assays are costly and tedious.

Advances in next-generation sequencing enabled genome-wide bioinformatic prediction of 200–500 bp CREs bound by TFs involved in cell fate specification and/or undergoing epigenomic remodeling of DNA methylation, histone modifications, and chromatin accessibility. Massively parallel reporter assays (MPRAs), or CRE-seq, emerged out of the need to test the functional relevance of thousands of putative CREs in a high-throughput manner (White 2015) (Fig. 59.1).

MPRAs build upon basic reporter assays, but instead of relying on a readout of β -galactosidase or luciferase activity, each putative CRE is paired with several short DNA sequence barcodes that are transcribed with the reporter gene (Kwasnieski et al. 2012; Melnikov et al. 2012) (Fig. 59.1e). CREs are synthesized on a microarray (<200 bp) or captured from the genome by DNA hybridization (>200 bp) (Shen et al. 2016) and then cloned into a vector chosen based on the target cell type/tissue, upstream of a minimal promoter that drives a basal level of reporter expression. The CRE-seq library is delivered to desired cell types by transfection, electroporation, or transduction if a viral vector is chosen [e.g., adeno-associated virus (AAV)]. Finally, the regulatory activity of each CRE is quantified using its RNA:DNA ratio of corresponding barcodes from next-generation sequencing read counts (Fig. 59.1f).

The mouse retina was an ideal model for developing CRE-seq, due to ease of library delivery, particularly by electroporation of newborn retinal explants. During 8 days in culture, the retinal explants continue neurogenesis and cell differentiation (Montana et al. 2011), providing an ex vivo system for efficient and reproducible delivery of plasmid CRE libraries. The plasmid DNA remains episomal, thus avoiding integration site effects. Further, transfected cells are mostly proliferating rod precursors, leading to highly reproducible reporter expression in a uniform cell type, thus allowing robust analysis of CRE activity (Montana et al. 2011).

MPRAs have been applied to two main goals: (1) identifying functional CREs genome-wide and (2) understanding the *cis*-regulatory grammar encoded within functional CREs. Here, we review several studies illustrating the use of MPRAs in the mouse retina and the knowledge gained about retinal *cis*-regulatory networks.

59.2 MPRAs to Identify Functional CREs Genome-Wide

59.2.1 Identification of Retina-Specific CREs and Mapping of Critical Regions for CRE Activity

Tissue-specific DNase hypersensitive sites (DHSs) have been hypothesized to have tissue-specific *cis*-regulatory activity (Natarajan et al. 2012; Heinz et al. 2015). To investigate the *cis*-regulatory activity of tissue-specific DHSs in the retina and cerebral cortex, ~3500 CREs (400–500 bp long) tiling central 300 bp DHSs were captured and cloned into CRE-seq vectors and then delivered into mouse retina explants by electroporation or into the mouse cortex by AAV in vivo (Shen et al. 2016). CREs with retina- or cortex-specific *cis*-regulatory activity were identified; further, “truncation mutations” generated by tiled CREs defined smaller critical regions for activity in some CREs.

59.2.2 Examining the Function of Cone-Rod Homeobox (CRX) Genome-Wide Binding Sites

CRX is a key driver of photoreceptor cell fate and maintenance. Its 7 bp binding motif occurs frequently across the genome, yet only a subset of those are bound by CRX as determined by ChIP-seq (Corbo et al. 2010). To determine local sequence determinants that distinguish CRX-bound sites from spurious motifs, White et al. designed a synthetic CRE-seq library containing 1300 unique CRX-bound genomic regions and 3000 control sequences including mutated CRX-bound sites, unbound CRX motifs, and scrambled controls representing a background level of CRE activity (White et al. 2013). Upon electroporation of the library into newborn mouse retinas, only CRX-bound regions could activate or repress transcription beyond background levels, suggesting that *cis*-regulatory activity of CRX-bound vs. unbound sequences was encoded in the local sequence flanking each CRX motif. Higher GC content of CRX-bound relative to unbound sequences was found to be the strongest distinguishing feature.

To examine if the *cis*-regulatory response of CRX-bound regions depends on CRX, the same library was assayed in both wild-type and *Crx*^{-/-} mouse retinas (White et al. 2016). The CRX-bound CREs displayed activity changes in response to the presence or absence of CRX protein. In contrast, regardless of the presence of CRX, neither unbound sequences nor CREs with mutated CRX motifs displayed *cis*-regulatory activity beyond background levels. This indicated that CRX-dependent *cis*-regulatory activity was encoded in the CRX-bound sequences. These studies illustrated the use of MPRAs to identify functional CREs and sequence features that predict *cis*-regulatory activity.

59.3 MPRAs to Decipher *Cis*-Regulatory Grammar

59.3.1 Identifying Critical Regulatory Sequences within a Known Rho CRE

A 52 bp CRE (*RhoCRE3*) located upstream of the rhodopsin (*Rho*) transcription start site is known to be sufficient to drive high levels of expression in rods (Lem et al. 1991; Zack et al. 1991). To dissect regulatory sequences within *RhoCRE3*, a synthetic CRE-seq library of >1000 wild-type and single- and double-nucleotide mutated variants of *RhoCRE3* was constructed and electroporated into newborn mouse retinas (Kwasnieski et al. 2012).

86% of single-nucleotide substitutions within *RhoCRE3* significantly affected regulatory activity. Mutations in binding motifs for known TFs, CRX, and NRL (neural retina leucine zipper) accounted for some activity changes. However, nonadditive activity changes caused by mutations in both CRX and NRL motifs highlighted the importance of TF synergistic interactions for proper *cis*-regulation of *Rho* expression, consistent with earlier studies demonstrating that CRX/NRL cooperative binding at genomic regions might be responsible (Hao et al. 2012) for synergistic CRX and NRL interactions in the retina (Rehemtulla et al. 1996; Mears et al. 2001; Corbo et al. 2007).

59.3.2 Investigating the Combinatorial Effect of TF Binding on Cis-Regulatory Activity

To determine the contribution of NRL binding sites to the expression level of specific CRX targets, a synthetic CRE-seq library was designed to represent varying numbers of CRX and NRL sites in enhancer and promoter sequences (White et al. 2016). Optimal reporter activity required a moderate number of CRX sites in both enhancers and promoters, but increasing CRX binding affinity (more binding motifs) increased the likelihood that the CRE would be repressive. However, adding one NRL site was sufficient to override the repression and turn the CRE into an activator. These findings suggested that varying the binding affinity and number of binding sites for CRX are sufficient to modulate its regulatory activity, while adding to the evidence that NRL and CRX bind cooperatively to modulate *cis*-regulatory activity.

Similarly, since CRX can act as a monomer or dimer, a MPRA was designed to investigate the regulatory activity of CRX monomeric vs. dimeric sites within CRX-responsive CREs (Hughes et al. 2018). A synthetic CRE-seq library representing saturation mutagenesis of monomeric and dimeric CRX binding sites revealed that dimeric sites encode stronger regulatory activity than monomeric sites. Moreover, mutating either half-site of a dimeric site abolished its activity, demonstrating the necessity of cooperation between half-sites for dimeric site activity. Additionally, altering spacing and nucleotide content between CRX dimeric half-sites within CREs showed that a strict 3 nucleotide space between half-sites was required for activity and CCG nucleotide content of the spacer was optimal.

59.4 Advantages and Limitations of MPRA

The studies above demonstrate the use of MPRA for elucidating the function of many CREs simultaneously. Comparisons of relative CRE activity within a CRE-seq library are inherently controlled since the experimental conditions are identical. But a major advantage is the modular nature of CRE-seq library construction, where the CRE, promoter, reporter, and delivery vector can each be changed to achieve specific gene expression.

Currently, most MPRA in mouse retinas rely on electroporation of plasmid DNA into newborn retina explants, resulting in transient expression of the CRE-seq library primarily in rods (Montana et al. 2011). In vivo AAV delivery will allow CRE-seq to be performed in tissues not amenable to transfection and increases the duration over which CRE activity can be studied (Shen et al. 2016). Further, different AAV serotypes can deliver the library to specific cells, allowing *cis*-regulatory networks of minor cell types to be studied. However, inserts packaged into AAV are limited to ~4.7 kb (Wu et al. 2010). Lentiviruses have a larger

capacity of ~8 kb (Kumar et al. 2001); but since its cargo is stably integrated into the host genome, CRE activity would be affected by *cis*-regulatory effects of sequences surrounding the integration site. On the other hand, while episomal expression allows CRE activity to be assessed independently of genomic context, integration into the genome could be advantageous if regulatory effects of the broader chromatin environment on CRE activity are of interest.

59.5 Future Applications of MPRA

More studies dissecting *cis*-regulatory grammar of noncoding DNA sequences will improve annotations of noncoding DNA function and allow better predictions of CRE activity for the purposes of synthetic biology. One can envision the use of MPRA in the development of gene therapies for retinopathies. MPRA could screen many combinations of enhancer, promoter, and vector sequences, to identify those able to drive expression of the gene product at appropriate levels in the correct cell type; saturation mutagenesis of these sequences might further optimize expression of the gene product.

In conclusion, while the MPRA was designed to study *cis*-regulatory function of noncoding DNA elements, its utility also potentially extends into accelerating diagnosis and design of therapies for genetic diseases in the retina and beyond.

Acknowledgments

This work is supported by NIH grants R01 EY012543 and EY025272 (to SC), postdoctoral fellowship T32 EY013360 (to WUSM), and RD2018 travel award (to IYO).

Abbreviations

CRE	<i>Cis</i> -regulatory element
CRX	Cone-rod homeobox
IRD	Inherited retinal degeneration
MPRA	Massively parallel reporter assay
NRL	Neural retina leucine zipper
TF	Transcription factor

References

- Corbo JC, Lawrence KA, Karlstetter M et al. (2010) CRX ChIP-seq reveals the *cis*-regulatory architecture of mouse photoreceptors. *Genome Res* 20:1512–1525 [PubMed: 20693478]
- Corbo JC, Myers CA, Lawrence KA et al. (2007) A typology of photoreceptor gene expression patterns in the mouse. *Proc Natl Acad Sci* 104:12069 [PubMed: 17620597]
- Hao H, Kim DS, Klocke B et al. (2012) Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL Targetome analysis Barsh GS, ed. *PLoS Genet* 8:e1002649 [PubMed: 22511886]
- Heinz S, Romanoski CE, Benner C et al. (2015) The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 16:144 [PubMed: 25650801]

- Hughes AEO, Myers CA, Corbo JC (2018) A massively parallel reporter assay reveals context-dependent activity of homeodomain binding sites in vivo. *Genome Res* 28(10):1520–1531 [PubMed: 30158147]
- Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76:8–32 [PubMed: 15549674]
- Kumar M, Keller B, Makalou N et al. (2001) Systematic determination of the packaging limit of lentiviral vectors. *Hum Gene Ther* 12:1893 [PubMed: 11589831]
- Kwasnieski JC, Mogno I, Myers CA et al. (2012) Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci U S A* 109:19498–19503 [PubMed: 23129659]
- Lem J, Applebury ML, Falk JD et al. (1991) Tissue-specific and developmental regulation of rod opsin chimeric genes in transgenic mice. *Neuron* 6:201 [PubMed: 1825171]
- Mears AJ, Kondo M, Swain PK et al. (2001) Nrl is required for rod photoreceptor development. *Nat Genet* 29:447–452 [PubMed: 11694879]
- Melnikov A, Murugan A, Zhang X et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 30:271–277 [PubMed: 22371084]
- Montana CL, Myers CA, Corbo JC (2011) Quantifying the activity of cis-regulatory elements in the mouse retina by explant electroporation. *J Vis Exp*:e2821–e2821
- Natarajan A, Yardimci GG, Sheffield NC et al. (2012) Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res* 22:1711 [PubMed: 22955983]
- Rehemtulla A, Warwar R, Kumar R et al. (1996) The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc Natl Acad Sci U S A* 93:191–195 [PubMed: 8552602]
- Shen SQ, Myers CA, Hughes AEO et al. (2016) Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res* 26:238–255 [PubMed: 26576614]
- Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 15:272–286 [PubMed: 24614317]
- Siemiatkowska AM, Collin RWJ, den Hollander AI et al. (2014) Genomic approaches for the discovery of genes mutated in inherited retinal degeneration. *Cold Spring Harb Perspect Med* 4
- Visel A, Bristow J, Pennacchio LA (2007) Enhancer identification through comparative genomics. *Semin Cell Dev Biol* 18:140–152 [PubMed: 17276707]
- White M, Myers C, Corbo J et al. (2013) Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci U S A* 109:11952–11957
- White MA (2015) Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics* 106:165–170 [PubMed: 26072432]
- White MA, Kwasnieski JC, Myers CA et al. (2016) A simple grammar defines activating and repressing cis-regulatory elements in photoreceptors. *Cell Rep* 17:1247–1254 [PubMed: 27783940]
- Wu Z, Yang H, Colosi P (2010) Effect of genome size on AAV vector packaging. *Mol Ther* 18:80–86 [PubMed: 19904234]
- Zack DJ, Bennett J, Wang Y et al. (1991) Unusual topography of bovine rhodopsin promoter-lacZ fusion gene expression in transgenic mouse retinas. *Neuron* 6:187 [PubMed: 1899580]

Fig. 59.1.

MPRA/CRE-seq design. **(a)** Putative CREs identified using epigenomic data such as TF binding, histone modifications, or chromatin accessibility. CREs synthesized or captured from the native genome can represent **(b)** genome-wide regions to be screened for regulatory function, or **(c)** a tiled "truncation" series or **(d)** saturation mutagenesis to define critical motifs for CRE activity. **(e)** CREs are uniquely barcoded, cloned into reporter vectors, and delivered to target cells/tissues. **(f)** Upon RNA/DNA isolation and sequencing of barcodes, the regulatory activity of each CRE is calculated using the ratio of corresponding RNA:DNA barcode counts. **(g)** Representative distribution of CRE activity within a CRE-seq library