A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors

Highlights

- CRX acts directly as both a repressor and activator in rod photoreceptors
- Activation and repression encoded by different affinity for CRX
- A binding site for the transcription factor NRL overrides repression
- Designed cis-regulatory elements recapitulate the behavior of genomic CRX targets

Authors

Michael A. White, Jamie C. Kwasnieski, Connie A. Myers, Susan Q. Shen, Joseph C. Corbo, Barak A. Cohen

Correspondence

cohen@genetics.wustl.edu

In Brief

Transcription factors often play different regulatory roles in the same cell. White et al. show how transcriptional activation and repression are encoded in regulatory DNA by the number and affinity of binding sites for the transcription factor CRX, enabling CRX to act as both repressor and activator in rod photoreceptors.
A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors

Michael A. White,1,3 Jamie C. Kwasnieski,1,3,4 Connie A. Myers,2 Susan Q. Shen,2 Joseph C. Corbo,2 and Barak A. Cohen1,5,*

1Department of Genetics and Center for Genome Sciences & Systems Biology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA
2Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA
3Co-first author
4Present address: Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
5Lead Contact

*Correspondence: cohen@genetics.wustl.edu

http://dx.doi.org/10.1016/j.celrep.2016.09.066

SUMMARY

Transcription factors often activate and repress different target genes in the same cell. How activation and repression are encoded by different arrangements of transcription factor binding sites in cis-regulatory elements is poorly understood. We investigated how sites for the transcription factor CRX encode both activation and repression in photoreceptors by assaying thousands of genomic and synthetic cis-regulatory elements in wild-type and Crx<sup>−/−</sup> retinas. We found that sequences with high affinity for CRX repress transcription, whereas sequences with lower affinity activate. This rule is modified by a cooperative interaction between CRX sites and sites for the transcription factor NRL, which overrides the repressive effect of high affinity for CRX. Our results show how simple rearrangements of transcription factor binding sites encode qualitatively different responses to a single transcription factor and explain how CRX plays multiple cis-regulatory roles in the same cell.

INTRODUCTION

A single transcription factor (TF) often plays multiple regulatory roles in the same cell by either activating or repressing different target genes (Alexandre and Vincent, 2003; Iype et al., 2004; Méthot and Basler, 1999; Parker et al., 2011). Such dual-function TFs occur in organisms from bacteria to mammals (Martínez-Montañés et al., 2013; Pompeani et al., 2008; Liu et al., 2014; Rachmin et al., 2015; Sánchez-Tilló et al., 2015). Because activation and repression occur in the same cell, the response of a target gene to a TF must be encoded in cis-regulatory elements by the specific arrangement, number, affinity, and identity of TF binding sites (Levo and Segal, 2014).

The ability of a single TF to both activate and repress is critical in the mammalian retina. The homeodomain TF CRX maintains the cell fate of rod and cone photoreceptors by activating and repressing key photoreceptor genes (Chen et al., 1997; Freund et al., 1997; Furukawa et al., 1997). In rods, which comprise more than 70% of the mouse retina (Jeon et al., 1998), CRX directly activates rod-specific genes, while repressing, directly or indirectly, cone-specific genes (Hsiau et al., 2007; Peng et al., 2005; Pittler et al., 2004). The TF binding sites and other sequence features that distinguish CRX-responsive cis-regulatory elements that activate from those that repress are unknown.

CRX interacts with the rod-specific leucine-zipper TF NRL, which is required to specify rod cell fate and suppress cone cell fate (Mears et al., 2001; Rehemtulla et al., 1996). NRL is necessary to regulate many activated and repressed CRX target genes in rods (Corbo et al., 2007), and NRL frequently binds genomic regions also bound by CRX (Hao et al., 2012). Thus, NRL binding sites likely contribute to the specific expression of CRX targets.

We sought to discover how cis-regulatory elements encode different transcriptional responses to CRX. We used massively parallel reporter assays to measure the activity of large numbers of genomic and synthetic cis-regulatory elements in wild-type and Crx<sup>−/−</sup> retinas. We found that different combinations of CRX and NRL sites distinguish activating from repressing sequences, revealing how CRX evokes qualitatively different transcriptional activities from different cis-regulatory elements in the same cell.

RESULTS

Genomic Sequences with High Affinity for CRX Repress Transcription

We previously reported that many genomic regions bound in vivo by CRX repress transcription in wild-type photoreceptors, when placed 5′ of the murine Rho proximal promoter (White et al., 2013). To determine whether CRX acts directly as a repressor at these sequences, we tested whether repression depends on CRX. We used cis-regulatory element analysis by sequencing (CRE-seq) (Kwasnieski et al., 2012; Mogno et al., 2013; White et al., 2013; Kwasnieski et al., 2014; Fiore and Cohen, 2016), a massively parallel reporter assay (Arnold et al., 2013; Melnikov
et al., 2012; Patwardhan et al., 2012; Shen et al., 2016; White, 2015), to measure the cis-regulatory activity of CRX-bound genomic sequences in wild-type and Crx−/− retinas. We assayed a library of 4,300 barcoded plasmid reporter genes (White et al., 2013), which included 865 short (84-bp) sequences taken from the centers of CRX chromatin immunoprecipitation sequencing (ChIP-seq) peaks (Corbo et al., 2010) and placed upstream of the Rho promoter. The library also included two sets of controls: (1) a mutant version of each ChIP peak sequence, in which all CRX motifs were abolished; and (2) randomized, negative control sequences, produced by scrambling the genomic sequences. The scrambled sequences create an empirical null distribution of reporter activity against which we defined activation and repression.

Repression by CRX-bound sequences required both CRX motifs and CRX protein. When assayed in wild-type retina, 22.1% of sequences drove strong reporter activation (expression above the 95th percentile of the scrambled negative controls), whereas 13.7% strongly repressed transcription (below the 5th percentile of the controls) (Figure 1A, left panel). This activity was lost when CRX motifs were abolished, demonstrating that both activation and repression depend on CRX sites (Figure 1A, right panel). When the same sequences were assayed in Crx−/− retina, we observed a striking loss of repression: only 3.74% of the ChIP peak sequences repressed reporter activity below the negative controls (Figure 1B, left panel). We thus conclude that CRX acts directly as a repressor at many of its genomic binding sites.

Despite the absence of CRX protein, many CRX ChIP peak sequences (24.1%) strongly activated in Crx−/− retina (Figure 1B, left panel), and this activity required intact CRX sites (Figure 1B, right panel). Many CRX ChIP peaks gained activity in Crx−/− retina, including 65% of sequences that repressed in wild-type retina (Figure 1C, red points), whereas few CRX ChIP peaks lost activity (Figure 1C, blue points). These results indicate that a second TF activates but does not repress through CRX binding sites. This TF is unknown; however, a possible candidate is OTX2, which recognizes a binding motif similar to that of CRX (Chatelain et al., 2006), which is highly expressed in rods in the early postnatal retina (Montana et al., 2011), and which can bind and activate some CRX targets (Koike et al., 2007; Samuel et al., 2014). Our results show that CRX binding sites mediate both activation and repression in a manner that depends on the identity of the TF that binds them.

How do CRX ChIP peak sequences encode qualitatively different cis-regulatory responses to CRX? We hypothesized that sites for additional TFs might define activating or repressive CRX-bound sequences. We searched for known or de novo motifs that were enriched in activating or repressing CRX ChIP-seq peaks, using several motif discovery tools (see Experimental Procedures); however, no additional motifs were enriched in either class of sequences. This includes the NRL motif; despite

Figure 1. CRX Is Necessary for Repression

A and B) Distribution of reporter expression from the Rho promoter by CRX ChIP peak sequences (blue), compared to scrambled negative controls (gray), in wild-type (A) or Crx−/− (B) retina. CRX motifs were abolished in mutant ChIP peak sequences (right panels). Dashed lines show 5th and 95th percentiles of scrambled controls used to define strongly repressing and activating ChIP peaks. Percentages (blue) indicate the fraction of CRX ChIP peaks below the 5th and above the 95th percentile of scrambled controls.

C) Relationship between reporter expression by wild-type CRX ChIP peak sequences in wild-type (x axis) and Crx−/− (y axis) retina. Sequences gained (red), lost (blue), or showed no significant change (black) in expression in Crx−/− retina.
were more likely to repress, whereas sequences with lower affinity scores were sequences less active than sequences with lower affinity scores, suggesting that, consistent with our model, the average CRX affinity required to repress transcription increased in the presence of the TATA promoter. In addition, most (55%) genomic sequences with the lowest CRX affinity scores were inactive. These results support the hypothesis that robust activation requires cis-regulatory elements with an optimal, moderate number of CRX sites.

CRX Sites Sufficient for Activation and Repression by Synthetic cis-Regulatory Elements

To directly test the hypothesis that activation and repression are encoded by the number and affinity of CRX sites, we turned to a simplified system of synthetic cis-regulatory elements (Gertz and Cohen, 2009; Gertz et al., 2009; Kwasnieski et al., 2012; Mogno et al., 2013; Sharon et al., 2012). Synthetic elements avoid the heterogeneity of genomic sequences, allowing us to directly test the contribution of CRX binding site number and affinity to activation and repression. We constructed a barcoded reporter gene library of synthetic cis-regulatory elements composed of combinations of three different CRX sites (high, moderate, or low affinity). Because CRX and NRL are known to activate synergistically at some CRX target genes, we also included NRL sites in some sequences. The library contained 1,290 designed sequences comprised of one to four binding sites, with sites occurring in either the forward or reverse orientation. As with the library of genomic CRX ChIP peaks, synthetic cis-regulatory elements were placed upstream of the Rho proximal promoter. Because these sequences contain only binding sites for CRX or NRL, they are a direct test of the ability of these sites to encode activation versus repression.

Our results confirm that CRX binding sites are sufficient to encode both activation and repression of the Rho promoter.
Considering sequences with CRX sites only, we found that synthetic cis-regulatory elements with lower affinity scores nearly always activated reporter expression above basal activity of the Rho promoter, whereas sequences with higher CRX affinity scores were more likely to repress (Figure 3A, left panel). Most synthetic elements with the highest affinity scores were repressive. However, the addition of a single NRL site abolished repression by sequences with high CRX affinity scores and transformed them into strongly activating cis-regulatory elements (Figure 3A, right panel). This suggests that the cooperative interaction between CRX and NRL overrides the repressive effect of high CRX affinity, converting repressive elements into strong activators.

To confirm that repression by synthetic cis-regulatory elements depends on CRX, we assayed the library in Crx−/− retina. As with the genomic sequences, synthetic cis-regulatory elements with high CRX affinity scores were de-repressed, confirming that CRX protein is necessary for repression (Figure S2). Additionally, strongly active synthetic cis-regulatory elements remained active in Crx−/− retina (Figure S2), further supporting the existence of a second TF that is able to activate but not repress from CRX binding sites in the absence of CRX protein.

Eliminating Promoter CRX Sites Reduces Repression by Synthetic Sequences

We further tested the hypothesis that high numbers of CRX sites in synthetic sequences repress, by replacing the Rho promoter with the Hsp68 promoter, which contains no CRX sites. We chose the Hsp68 promoter, rather than the minimal TATA promoter, because Hsp68 completely lacks CRX sites and has some autonomous basal activity, which allows us to measure both activation and repression. In agreement with the prediction of our model, synthetic cis-regulatory elements that repressed the Rho promoter became strongly activating with the Hsp68 promoter (Figure 3B, left panel). In contrast to results with the Rho promoter, activation of the Hsp68 promoter increased with increasing CRX affinity scores throughout the entire observed range. Because Hsp68 has no CRX sites, this result indicates that few synthetic cis-regulatory elements have a sufficiently high affinity for CRX to cause repression on their own in the absence of promoter CRX sites. Addition of an NRL site still synergistically activated expression with CRX sites, driving stronger activation than the corresponding cis-regulatory elements without an NRL site (Figure 3B, right panel; compare with left panel).

Taken together, these results confirm that CRX sites alone are sufficient to encode activation and repression. The results are concordant with the patterns of expression directed by the genomic sequences, and support the hypothesis that CRX-responsive cis-regulatory elements are governed by a simple regulatory rule. Optimal activation is achieved by a moderate affinity for CRX, whereas higher CRX affinity leads to repression. This rule is modified by the presence of an NRL site, which causes sequences with higher affinity for CRX to activate rather than repress.

Number and Affinity of CRX Sites Govern Activation and Repression in the Genome

We tested whether our proposed cis-regulatory grammar accounts for activation and repression by CRX-responsive elements within the larger sequence context of the genome. First, we computed the CRX affinity scores of all CRX-bound genomic...
regions near genes activated or repressed by CRX. Although TF binding sites do not always regulate the closest gene, we took the photoreceptor genes nearest CRX ChIP peaks as an approximation of genes regulated by CRX (Corbo et al., 2010). CRX-bound regions near repressed genes were more likely to have higher CRX affinity scores than CRX-bound regions near activated genes (Figure 4A), consistent with the prediction of our proposed cis-regulatory grammar. Some of these CRX-bound regions were also included in our reporter gene assay (near repressed genes, n = 24; near activated genes, n = 78). CRX-bound regions near repressed photoreceptor genes were more likely to show loss of reporter gene repression in Crx−/− retina, whereas sequences near activated genes were more likely to lose reporter activity in Crx−/− retina (p = 0.04, Wilcoxon test), further supporting the genomic validity of our reporter assay results.

Finally, to test the hypothesis that an interaction between CRX and NRL drives stronger activation, we asked whether CRX and NRL co-bind near highly expressed photoreceptor genes. We identified CRX ChIP-seq peaks (Corbo et al., 2010) that overlap NRL ChIP-seq peaks (Hao et al., 2012), and examined the expression of nearby photoreceptor genes, as reported in a previous RNA-sequencing study (Brooks et al., 2011). Photoreceptor genes near regions co-bound by CRX and NRL tended to express more highly than genes near regions bound only by CRX (Figure 4B). We also found that co-bound regions have higher CRX affinity scores than sequences bound by CRX alone (Figure 4C). These results are consistent with the hypothesis that an interaction between CRX and NRL overrides the repressive effects of high CRX affinity. Taken together, these orthogonal genomic data support the validity of our hypothesized cis-regulatory grammar within the full context of the genome.

**DISCUSSION**

Many TFs are bi-functional, acting as repressors or activators depending on cellular state and the sequence context of their cis-regulatory targets. In some cases, this dual function results from post-translational modifications (Méthot and Basler, 1999; Parker et al., 2011), whereas in others, binding sites for additional factors encode different activities at different cis-regulatory elements (Alexandre and Vincent, 2003; Martínez-Montañés et al., 2013; Pompeani et al., 2008; Rachmin et al., 2015; Sánchez-Tilló et al., 2015). We have shown here that variation in the number and affinity of binding sites for only a single TF is sufficient to encode activation versus repression. Our data suggest a “Goldilocks” hypothesis in which robust activation requires cis-regulatory elements with an optimal affinity for CRX; sequences with too few CRX sites fail to activate, whereas too many CRX sites push the system into a repressive regime, except when an NRL site is present. We note that this cis-regulatory grammar does not suffice to distinguish genuine CRX binding sites from background genomic sequence. We find that multiple low-to-moderate affinity CRX sites drive robust activation; however, it is unknown how functional CRX sites are distinguished from spurious clusters of low-affinity motifs (White et al., 2013).

Our results indicate that the optimal CRX affinity required to activate transcription depends on CRX motifs in both distal enhancers and proximal promoters. Many well-characterized photoreceptor genes contain promoter CRX binding sites (Corbo et al., 2010), and our model predicts that these sites contribute the CRX affinity-encoded regulation of photoreceptor genes. The requirement for an optimal enhancer/promoter CRX affinity likely operates with additional enhancer TF binding sites that recruit trans factors that are biochemically compatible with the core transcriptional machinery present at different promoters (Zabidi et al., 2015).

Multiple studies show that homotypic clusters of sites for a single TF can drive lower expression than heterotypic clusters of sites for multiple TFs, suggesting that cooperative interactions between different TFs are required for optimal activation (Fiore and Cohen, 2016; Levo and Segal, 2014; Smith et al., 2013). Consistent with this, our data show that cooperativity between CRX and NRL drives the highest expression and overcomes the repressive effects of homotypic clusters of CRX sites. The mechanism underlying affinity-dependent repression by CRX,
a well-established transcriptional activator (Peng et al., 2005; Pittler et al., 2004), is unknown. One model is that binding of one TF molecule inhibits the binding of additional TF molecules in the presence of multiple, closely spaced binding sites (Fiore and Cohen, 2016; Levo et al., 2015). cis-Regulatory elements with multiple distal CRX sites may thus prevent CRX binding to critical promoter sites, causing repression below basal levels. An alternate model is that high CRX occupancy occludes binding of additional transcriptional co-factors at the promoter, which is relieved by NRL or other cooperatively interacting TFs that recruit those additional co-factors. Deciphering the mechanism of this affinity-dependent switch between activation and repression should be an important goal of future studies.

EXPERIMENTAL PROCEDURES

Construction of Barcoded Reporter Gene Libraries CRX ChIP Peak Reporter Libraries

We built barcoded reporter libraries of CRX ChIP peak sequences with the Rho proximal promoter as described (White et al., 2013). Briefly, library sequences were synthesized as barcoded oligonucleotides by Agilent and cloned into a vector backbone. The rod-specific Rho promoter and the DsRed gene were cloned between the ChIP peak sequence and the barcode. In CRE-seq, DsRed serves only as a spacer between the promoter and transcribed barcode and its fluorescence is not measured. Each cis-regulatory sequence in the library was represented by three unique barcodes. For the promoter replacement experiment, the Rho promoter was replaced with a minimal TATA promoter consisting of the region −36 to +79 around the TATA box of the bovine Rho promoter, described previously (Corbo et al., 2010; Hsiau et al., 2007).

Synthetic cis-Regulatory Element Libraries

The library of synthetic elements contained 1,290 sequences composed of combinations of CRX and NRL sites that were up to four TF sites in length. We used two CRX sites of differing affinity, taken from the murine Rho promoter (Kwasnieski et al., 2012); the high-affinity consensus sequence (CTAATCCC) and a moderate-affinity site (CTAAGCCA). We also used a low-affinity CRX site (CTGATTCA), which we hypothesized is bound by CRX (CTAATCCC) and a moderate-affinity site (CTAAGCCA). We also used a low-affinity CRX site (CTGATTCA), which we hypothesized is bound by CRX

Calculation of Barcoded Reporter Gene Expression

As described previously (Kwasnieski et al., 2012), expression of each barcoded reporter gene was determined by the number of RNA reads of the corresponding barcode. To account for differences in barcode representation in the pooled library, RNA reads were normalized to DNA reads. RNA/DNA ratios were averaged over all barcodes for each element. Due to the expected lack of expression of many reporter constructs with the minimal TATA promoter (Figure S1), it was necessary to distinguish reliably low expression by weakly active reporter constructs from spurious detection of inactive reporter constructs (e.g., inactive reporters producing zero RNA reads in most replicates and many RNA reads in a single, outlier replicate). We thus applied a coefficient of variance (CV) threshold to CRE-seq results. We added a pseudo-count to the RNA reads for each barcode (to eliminate values of zero), and then calculated the CV. Following normalization by DNA reads, barcodes with a CV above an empirically determined threshold of 1.15 were discarded. This threshold eliminated outlier barcodes that produced no RNA signal in at least one-half of the replicates, but high RNA signal in a minority of replicates.

Retinal Explant Electroporation and CRE-Seq Assay

Electroporation into retinal explants and barcode RNA and DNA sequencing were performed as described previously (Kwasnieski et al., 2012; White et al., 2013). Retinas were harvested from newborn (postnatal day 0) C57BL/6 and Crx ∆/∆ mice as described (Hsiau et al., 2007). CD-1 mice were used for the library with the minimal TATA promoter for consistency with our previous experiments in this strain (White et al., 2013). Three or four electroporations were performed for each experiment. Reporter expression measurements in replicate electroporations were highly correlated (Pearson’s correlation coefficient between replicates > 0.95). Animal procedures were performed in accordance with a Washington University School of Medicine Animal Studies Committee-approved vertebrate animals protocol.

Calculation of Significantly Changed Reporter Expression in Crx ∆/∆ Retina

Significantly changed expression was determined by comparison with the empirical null distribution of the scrambled sequences. CRX ChIP peaks were considered to have significantly gained or lost activity in Crx ∆/∆ (red and blue points, Figure 1C) retina if change in activity was greater than that observed for 95% of the scrambled sequences.

Motive Analysis of Genomic Sequences

To search for de novo and known motifs in CRX ChIP peak sequences, we used the MEME suite (MEME [Bailey et al., 2006], DREME [Bailey, 2011], and MEME-ChIP [Machanick and Bailey, 2011]) and the web server version of k-mer SVM (Fietz-Brant et al., 2013). To directly identify occurrences of the canonical NRL motif, we used FIMO (Grant et al., 2011) and an NRL position-weight matrix (Jolma et al., 2013).

Calculation of CRX Affinity Scores

We calculated CRX affinity scores (previously referred to as “predicted occupancy”) as described (White et al., 2013). Unlike a motif-scoring approach based on a p value threshold (Grant et al., 2011), CRX affinity scores are a threshold-free measure of aggregate CRX affinity that considers both the number and affinity of CRX sites. These scores are obtained from the binding model described in Equation 1 of Zhao et al. (2009), using a μ parameter of 9 (White et al., 2013) and the CRX position weight matrix determined by Lee et al. (2010). Scores for the forward and reverse complement sequences were summed to produce a total score for each sequence. Cluster-Buster was run with the gap parameter set to 5 bp and the minimum reported cluster score set to 1.

Statistical Analysis

Two-sample comparisons of reporter gene data were performed using two-tailed Welch’s t test. For comparisons of multiple samples, Pearson’s chi-square test was performed. For two-sample comparisons of genomic data (Figure 4), the Wilcoxon rank sum test was performed. Data were considered statistically significant at p < 0.05.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and four databases and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.09.066.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

We thank Andrew Hughes for assistance with the analysis of the NRL ChIP-seq data. This work was supported by NIH grants HG006346 (to B.A.C. and J.C.C.) and GM092910 (to B.A.C.).

REFERENCES


Cell Reports 17, 1247–1254, October 25, 2016 1253
zipper (Nrl), a photoreceptor cell fate determinant. J. Biol. Chem. 286, 36921–36931.


1254 Cell Reports 17, 1247–1254, October 25, 2016