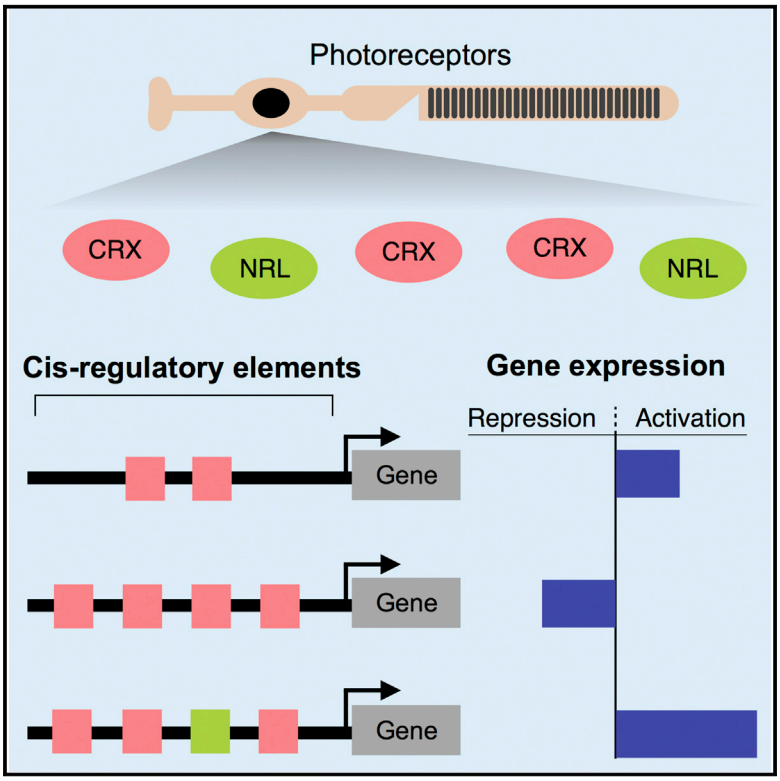
Report

Cell Reports

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

Graphical Abstract



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.

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

White et al., 2016, Cell Reports *17*, 1247–1254
October 25, 2016
http://dx.doi.org/10.1016/j.celrep.2016.09.066

CellPress

Cell Reports
Report

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. Cohen[1,5,*]

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

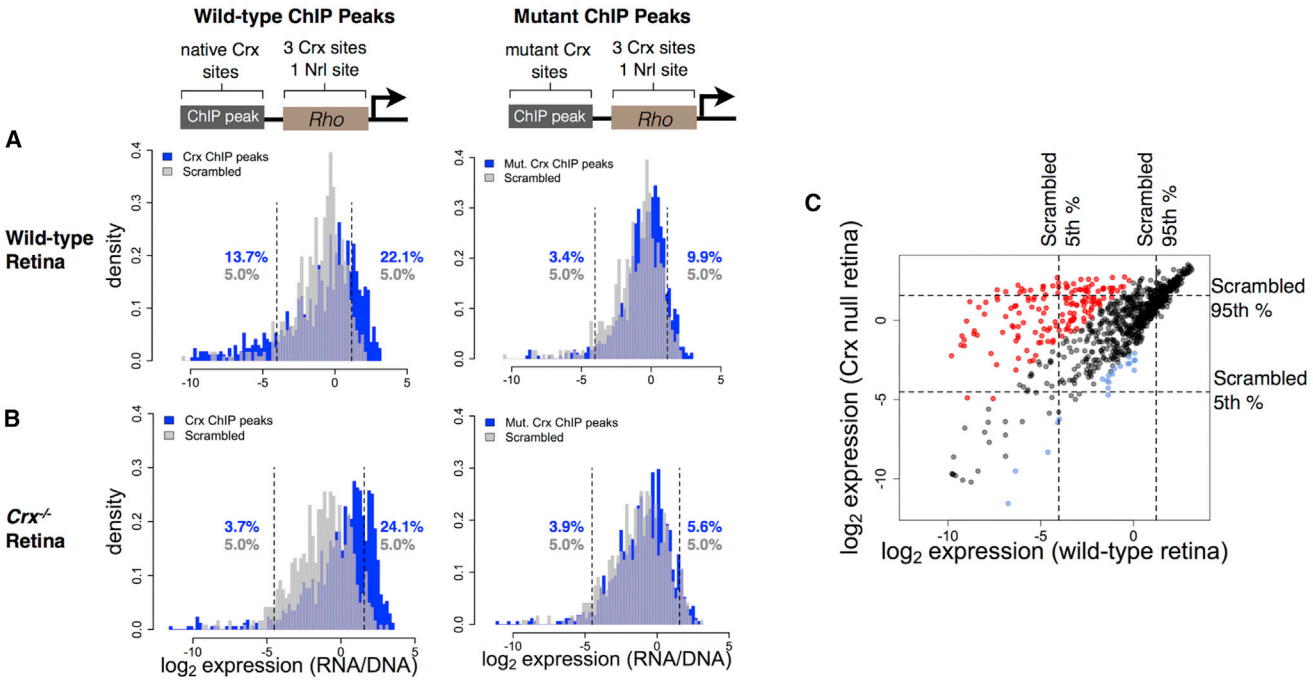


Figure 1. CRX Is Necessary for Repression

(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.

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

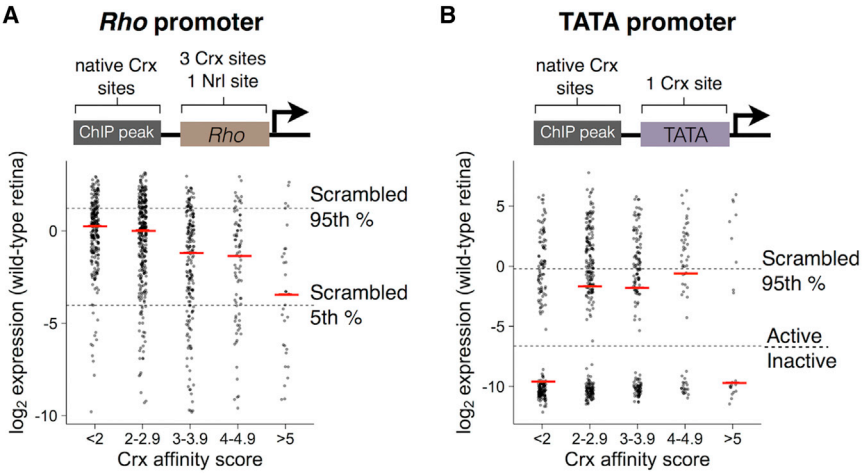


Figure 2. Activation and Repression Encoded by the Number of CRX Binding Sites

(A and B) Activation and repression of reporter expression (y axis) by CRX ChIP peak sequences in wild-type retina depend on number and affinity of CRX motifs (CRX affinity scores, x axis). Reducing the number of CRX sites in reporter constructs by replacing the *Rho* promoter (A) with a minimal TATA promoter (B) increases the optimal CRX affinity score of ChIP peak sequences necessary for robust activation. Red lines indicate median expression. Dashed lines indicate expression levels of 95th and 5th percentiles of the scrambled controls (*Rho*) and the threshold defining minimal reporter activity (TATA; see Experimental Procedures). See also Figure S1.

the frequent overlap of CRX and NRL binding in the genome (Hao et al., 2012), canonical NRL motifs occurred rarely in CRX ChIP peaks (present in fewer than 3% of sequences) and were not enriched relative to the scrambled controls.

The CRX motif itself strongly distinguished activating from repressing sequences. Specifically, the number and affinity of CRX motifs in repressive sequences were higher than in activating sequences. We computed an overall CRX affinity score for each genomic sequence using a threshold-free binding model that considers both the number and affinity of CRX sites (White et al., 2013; Zhao et al., 2009). These scores do not include the *Rho* proximal promoter present in the reporter constructs, which itself contains three CRX sites and an NRL site (Corbo et al., 2010; Kwasnieski et al., 2012), and has a CRX affinity score of 2.8. Genomic sequences with the highest CRX affinity scores were more likely to repress, whereas sequences with lower affinity scores were likely to activate (Figure 2A, $p = 1.3 \times 10^{-3}$, Pearson's chi-square test). To test the robustness of this finding, we scored the same sequences using Cluster-Buster, a probabilistic model-based tool to identify clusters of TF binding sites (Frith et al., 2003). Consistent with our CRX affinity scores, sequences that repressed contained higher scoring clusters of CRX motifs than sequences that activated (4.8 versus 4.1 mean cluster score, $p = 0.001$, Welch's two-sample t test). These results suggest that activation and repression by CRX ChIP peak sequences are partially encoded by the number and affinity of the CRX sites they contain: robust activation requires a moderate affinity for CRX, whereas high CRX affinity produces repression.

Removing CRX Sites Increases Output from Repressive Sequences

If repression is encoded by high overall CRX affinity, then reducing the number of CRX sites should switch repressive sequences to activating sequences. To test this counterintuitive prediction, we reduced the number of CRX sites in the reporter constructs by replacing the *Rho* promoter (with three CRX sites and a CRX affinity score of 2.8) with a short minimal promoter that contains a TATA box and only a single CRX site (CRX affinity score of 1.2). Unlike the *Rho* promoter, the minimal TATA promoter does not drive reporter expression on its own (Corbo et al., 2010), and thus only activating CRX ChIP peaks produce a reporter signal (Figure S1). In agreement with our prediction, sequences with higher CRX affinity scores were now more likely to drive higher expression than sequences with lower CRX affinity scores (Figure 2B, $p = 0.039$, Pearson's chi-square test; compare with Figure 2A). Only at the very highest CRX 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.

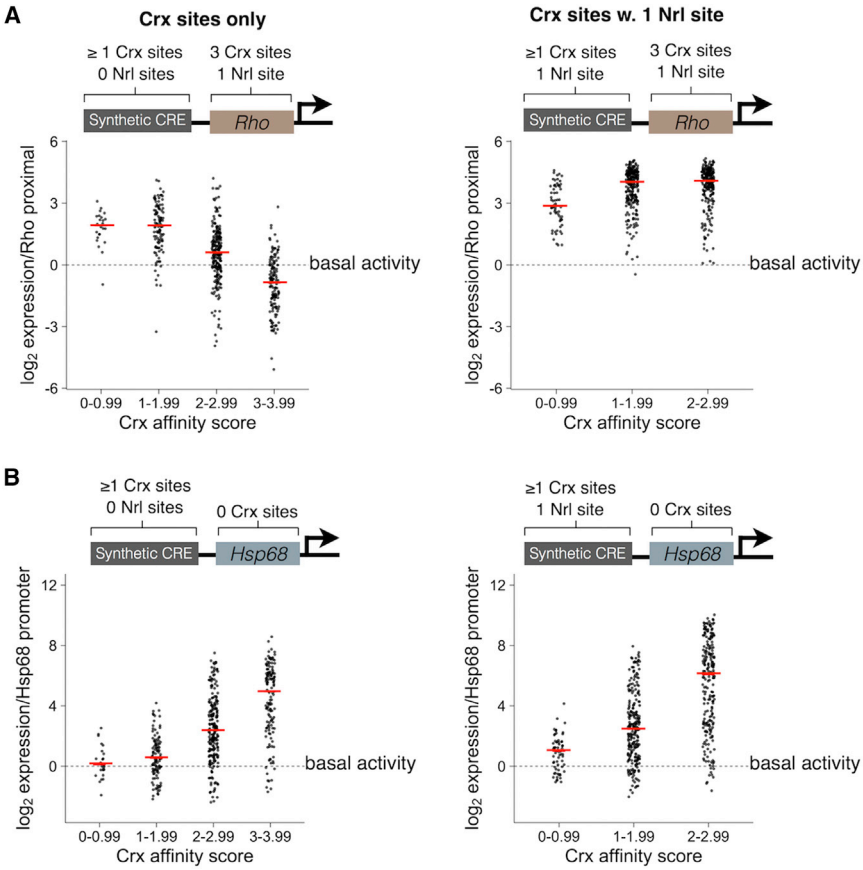


Figure 3. CRX and NRL Binding Sites Are Sufficient for Activation and Repression

For a Figure360 author presentation of Figure 3, see the figure online at http://dx.doi.org/10.1016/j.celrep.2016.09.066#mmc6

(A and B) Relationship between CRX affinity score (x axis) and reporter expression (y axis) of synthetic *cis*-regulatory elements (CREs) containing only CRX binding sites (left column) or CRX sites plus a single NRL site (right column), driving the (A) *Rho* proximal promoter (with CRX sites) or the (B) *Hsp68* promoter (without CRX sites). Reporter expression is normalized relative to basal activity of the promoter (indicated by the dashed line). Red lines indicate median expression. Synthetic *cis*-regulatory elements consist of up to four binding sites, and CRX affinity scores range from 0 to 3.99. See also Figure S2.

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

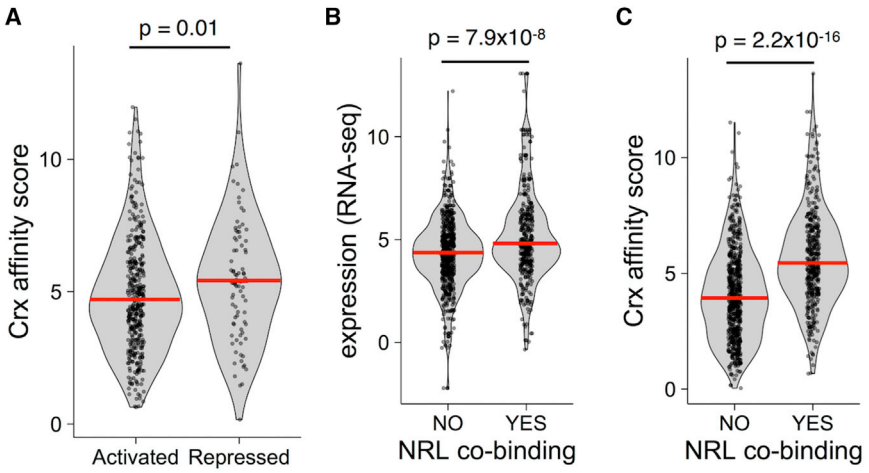


Figure 4. Simple Regulatory Grammar Governs Activation and Repression in the Genome

(A) Violin plot showing CRX affinity scores (y axis) of CRX ChIP peaks near genes that are activated or repressed by CRX ($p = 0.01$, Wilcoxon rank test; n = 388 activated genes, 85 repressed genes). CRX ChIP peaks near activated and repressed genes were identified in Corbo et al. (2010).

(B) Genes near genomic regions bound by both CRX and NRL are more highly expressed than genes near regions bound only by CRX ($p = 7.9 \times 10^{-8}$, Wilcoxon rank test; n = 379 peaks co-bound by NRL, 582 peaks not co-bound by NRL).

(C) CRX-bound regions that are co-bound by NRL have higher CRX affinity scores ($p = 2.2 \times 10^{-16}$, Wilcoxon rank test; n = 379 peaks co-bound by NRL, 582 peaks not co-bound by NRL). Red lines indicate median values. CRX and NRL binding data are from Corbo et al. (2010) and Hao et al. (2012); expression data are from Brooks et al. (2011).

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 based on evidence of competition (Kwasnieski et al., 2012). For the NRL site, we used the consensus sequence (Kataoka et al., 1994). Short constant buffer sequences were added to each site to maintain helical spacing when sites were combined. Using these four different TF binding sites, we generated synthetic CREs representing every possible combination of one, two, or three sites, and 715 of the possible 4,096 synthetic elements that are four sites long. Each sequence in the library was represented with five unique barcodes. These sequences were synthesized as custom oligonucleotides by Agilent and cloned as described above.

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 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 these 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.

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.

Motif 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 (Fletez-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 mu 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

Conceptualization, J.C.K., M.A.W., and B.A.C.; Methodology, J.C.K., B.A.C., J.C.C. and M.A.W.; Investigation, J.C.K., M.A.W., C.A.M., S.Q.S.; Formal Analysis, J.C.K and M.A.W.; Writing – Original Draft, M.A.W. and J.C.K.; Writing – Review & Editing, M.A.W., J.C.K., B.A.C, J.C.C., and S.Q.S.; Funding Acquisition, B.A.C. and J.C.C.

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.).

Received: May 5, 2016
Revised: August 6, 2016
Accepted: September 20, 2016
Published: October 25, 2016

REFERENCES

Alexandre, C., and Vincent, J.-P. (2003). Requirements for transcriptional repression and activation by Engrailed in *Drosophila* embryos. Development *130*, 729–739.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369–W373.

Brooks, M.J., Rajasimha, H.K., Roger, J.E., and Swaroop, A. (2011). Next-generation sequencing facilitates quantitative analysis of wild-type and $Nrl^{-/-}$ retinal transcriptomes. Mol. Vis. *17*, 3034–3054.

Chatelain, G., Fossat, N., Brun, G., and Lamonerie, T. (2006). Molecular dissection reveals decreased activity and not dominant negative effect in human OTX2 mutants. J. Mol. Med. *84*, 604–615.

Chen, S., Wang, Q.L., Nie, Z., Sun, H., Lennon, G., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Zack, D.J. (1997). Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. Neuron *19*, 1017–1030.

Corbo, J.C., Myers, C.A., Lawrence, K.A., Jadhav, A.P., and Cepko, C.L. (2007). A typology of photoreceptor gene expression patterns in the mouse. Proc. Natl. Acad. Sci. USA *104*, 12069–12074.

Corbo, J.C., Lawrence, K.A., Karlstetter, M., Myers, C.A., Abdelaziz, M., Dirkes, W., Weigelt, K., Seifert, M., Benes, V., Fritsche, L.G., et al. (2010). CRX ChIP-seq reveals the *cis*-regulatory architecture of mouse photoreceptors. Genome Res. *20*, 1512–1525.

Fiore, C., and Cohen, B.A. (2016). Interactions between pluripotency factors specify *cis*-regulation in embryonic stem cells. Genome Res. *26*, 778–786.

Fletez-Brant, C., Lee, D., McCallion, A.S., and Beer, M.A. (2013). kmer-SVM: a web server for identifying predictive regulatory sequence features in genomic data sets. Nucleic Acids Res. *41*, W544–W556.

Freund, C.L., Gregory-Evans, C.Y., Furukawa, T., Papaioannou, M., Looser, J., Ploder, L., Bellingham, J., Ng, D., Herbrick, J.A., Duncan, A., et al. (1997). Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. Cell *91*, 543–553.

Frith, M.C., Li, M.C., and Weng, Z. (2003). Cluster-Buster: Finding dense clusters of motifs in DNA sequences. Nucleic Acids Res. *31*, 3666–3668.

Furukawa, T., Morrow, E.M., and Cepko, C.L. (1997). Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. Cell *91*, 531–541.

Gertz, J., and Cohen, B.A. (2009). Environment-specific combinatorial *cis*-regulation in synthetic promoters. Mol. Syst. Biol. *5*, 244.

Gertz, J., Siggia, E.D., and Cohen, B.A. (2009). Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters. Nature *457*, 215–218.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Hao, H., Kim, D.S., Klocke, B., Johnson, K.R., Cui, K., Gotoh, N., Zang, C., Gregorski, J., Gieser, L., Peng, W., et al. (2012). Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL targetome analysis. PLoS Genet. *8*, e1002649.

Hsiau, T.H.-C., Diaconu, C., Myers, C.A., Lee, J., Cepko, C.L., and Corbo, J.C. (2007). The *cis*-regulatory logic of the mammalian photoreceptor transcriptional network. PLoS One *2*, e643.

Iype, T., Taylor, D.G., Ziesmann, S.M., Garmey, J.C., Watada, H., and Mirmira, R.G. (2004). The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: evidence for feedback activation of the nkx6.1 gene by Nkx6.1. Mol. Endocrinol. *18*, 1363–1375.

Jeon, C.J., Strettoi, E., and Masland, R.H. (1998). The major cell populations of the mouse retina. J. Neurosci. *18*, 8936–8946.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Kataoka, K., Noda, M., and Nishizawa, M. (1994). Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. Mol. Cell. Biol. *14*, 700–712.

Koike, C., Nishida, A., Ueno, S., Saito, H., Sanuki, R., Sato, S., Furukawa, A., Aizawa, S., Matsuo, I., Suzuki, N., et al. (2007). Functional roles of Otx2 transcription factor in postnatal mouse retinal development. Mol. Cell. Biol. *27*, 8318–8329.

Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2012). Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. Proc. Natl. Acad. Sci. USA *109*, 19498–19503.

Kwasnieski, J.C., Fiore, C., Chaudhari, H.G., and Cohen, B.A. (2014). High-throughput functional testing of ENCODE segmentation predictions. Genome Res. *24*, 1595–1602.

Lee, J., Myers, C.A., Williams, N., Abdelaziz, M., and Corbo, J.C. (2010). Quantitative fine-tuning of photoreceptor *cis*-regulatory elements through affinity modulation of transcription factor binding sites. Gene Ther. *17*, 1390–1399.

Levo, M., and Segal, E. (2014). In pursuit of design principles of regulatory sequences. Nat. Rev. Genet. *15*, 453–468.

Levo, M., Zalckvar, E., Sharon, E., Dantas Machado, A.C., Kalma, Y., Lotam-Pompan, M., Weinberger, A., Yakhini, Z., Rohs, R., and Segal, E. (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. *25*, 1018–1029.

Liu, Y.-R., Laghari, Z.A., Novoa, C.A., Hughes, J., Webster, J.R.M., Goodwin, P.E., Wheatley, S.P., and Scotting, P.J. (2014). Sox2 acts as a transcriptional repressor in neural stem cells. BMC Neurosci. *15*, 95.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

Martínez-Montañés, F., Rienzo, A., Poveda-Huertes, D., Pascual-Ahuir, A., and Proft, M. (2013). Activator and repressor functions of the Mot3 transcription factor in the osmostress response of *Saccharomyces cerevisiae*. Eukaryot. Cell *12*, 636–647.

Mears, A.J., Kondo, M., Swain, P.K., Takada, Y., Bush, R.A., Saunders, T.L., Sieving, P.A., and Swaroop, A. (2001). Nrl is required for rod photoreceptor development. Nat. Genet. *29*, 447–452.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30*, 271–277.

Méthot, N., and Basler, K. (1999). Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. Cell *96*, 819–831.

Mogno, I., Kwasnieski, J.C., and Cohen, B.A. (2013). Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. Genome Res. *23*, 1908–1915.

Montana, C.L., Lawrence, K.A., Williams, N.L., Tran, N.M., Peng, G.-H., Chen, S., and Corbo, J.C. (2011). Transcriptional regulation of neural retina leucine

zipper (Nrl), a photoreceptor cell fate determinant. J. Biol. Chem. *286*, 36921–36931.

Parker, D.S., White, M.A., Ramos, A.I., Cohen, B.A., and Barolo, S. (2011). The *cis*-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. Sci. Signal. *4*, ra38.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270.

Peng, G.-H., Ahmad, O., Ahmad, F., Liu, J., and Chen, S. (2005). The photoreceptor-specific nuclear receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus cone genes. Hum. Mol. Genet. *14*, 747–764.

Pittler, S.J., Zhang, Y., Chen, S., Mears, A.J., Zack, D.J., Ren, Z., Swain, P.K., Yao, S., Swaroop, A., and White, J.B. (2004). Functional analysis of the rod photoreceptor cGMP phosphodiesterase alpha-subunit gene promoter: Nrl and Crx are required for full transcriptional activity. J. Biol. Chem. *279*, 19800–19807.

Pompeani, A.J., Irgon, J.J., Berger, M.F., Bulyk, M.L., Wingreen, N.S., and Bassler, B.L. (2008). The *Vibrio harveyi* master quorum-sensing regulator, LuxR, a TetR-type protein is both an activator and a repressor: DNA recognition and binding specificity at target promoters. Mol. Microbiol. *70*, 76–88.

Rachmin, I., Amsalem, E., Golomb, E., Beeri, R., Gilon, D., Fang, P., Nechushtan, H., Kay, G., Guo, M., Yiqing, P.L., et al. (2015). FHL2 switches MITF from activator to repressor of Erbin expression during cardiac hypertrophy. Int. J. Cardiol. *195*, 85–94.

Rehemtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D.J., and Swaroop, A. (1996). The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. Proc. Natl. Acad. Sci. USA *93*, 191–195.

Samuel, A., Housset, M., Fant, B., and Lamonerie, T. (2014). Otx2 ChIP-seq reveals unique and redundant functions in the mature mouse retina. PLoS One *9*, e89110.

Sánchez-Tilló, E., de Barrios, O., Valls, E., Darling, D.S., Castells, A., and Postigo, A. (2015). ZEB1 and TCF4 reciprocally modulate their transcriptional activities to regulate Wnt target gene expression. Oncogene *34*, 5760–5770.

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. *30*, 521–530.

Shen, S.Q., Myers, C.A., Hughes, A.E.O., Byrne, L.C., Flannery, J.G., and Corbo, J.C. (2016). Massively parallel *cis*-regulatory analysis in the mammalian central nervous system. Genome Res. *26*, 238–255.

Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat. Genet. *45*, 1021–1028.

White, M.A. (2015). Understanding how *cis*-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. Genomics *106*, 165–170.

White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (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. USA *110*, 11952–11957.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559.

Zhao, Y., Granas, D., and Stormo, G.D. (2009). Inferring binding energies from selected binding sites. PLoS Comput. Biol. *5*, e1000590.