Enhancer transcription identifies *cis*-regulatory elements for photoreceptor cell types

Rangarajan D. Nadadur[1]*, Carlos Perez-Cervantes[1]*, Nicolas Lonfat[2]*, Linsin A. Smith[1]*, Andrew E. O. Hughes[3], Sui Wang[2,4], Joseph C. Corbo[3], Connie Cepko[2], Ivan P. Moskowitz[1]

*Contributed equally

Affiliations:
[1]Departments of Pediatrics, Pathology, and Human Genetics, University of Chicago, Chicago, IL, USA
[2] Departments of Genetics and Ophthalmology, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.
Electronic address: cepko@genetics.med.harvard.edu.
[3] Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA
[4] Current address: Department of Ophthalmology, Stanford University, Stanford, CA 94305, USA

Key words: cis-regulatory element (CRE), enhancer, transcription factor, non-coding transcriptome, non-coding RNA (ncRNA), long non-coding RNA (lncRNA), chromatin, gene regulatory network, photoreceptor, Nrl, Otx2, Onecut, rod, cone

Please send correspondence to:
Ivan Moskowitz, M.D., Ph.D.
Departments of Pediatrics, Pathology, and Human Genetics
The University of Chicago
900 East 57th Street, KCBD Room 5102
Chicago, Illinois 60637
Phone: 773/834-0462
imoskowitz@uchicago.edu

Abstract

Identification of the *cis*-regulatory elements (CREs) that regulate gene expression in specific cell types is critical for defining the gene regulatory networks (GRNs) that control normal physiology and disease states. We previously utilized non-coding RNA (ncRNA) profiling to define CREs that comprise a GRN in the adult mouse heart[1]. Here, we applied ncRNA profiling to the mouse retina in the presence and absence of *Nrl*, a rod photoreceptor-specific transcription factor required for rod versus cone photoreceptor cell fate. Differential expression of *Nrl*-dependent ncRNAs positively correlated with differential expression of *Nrl*-dependent local genes. Two distinct *Nrl*-dependent regulatory networks were discerned in parallel: *Nrl*-activated ncRNAs were enriched for accessible chromatin in rods but not cones whereas *Nrl*-repressed ncRNAs were enriched for accessible chromatin in cones but not rods. Furthermore, differential *Nrl*-dependent ncRNA expression levels quantitatively correlated with photoreceptor cell type-specific ATAC-seq read density. Direct assessment of *Nrl*-dependent ncRNA-defined loci identified functional cone photoreceptor CREs. This work supports differential ncRNA profiling as a platform for identifying context-specific regulatory elements and provides insight into the networks that define photoreceptor cell types.

Introduction

Identification of tissue and context specific *cis*-regulatory elements (CREs) is critical to defining the transcriptional networks that govern physiology and disease across biological contexts. The advent of high-throughput sequencing has allowed for genome-wide analysis of chromatin state as a proxy for regulatory elements[2]. The use of histone modifications, chromatin status and transcription factor (TF) occupancy to define gene regulation has been successful in many contexts[3-6]. However, these approaches lack both specificity and quantitative resolution of enhancer activity[2]. These observations indicate the need for complementary strategies for the identification of functional context-dependent enhancers.

A growing body of literature indicates that noncoding RNAs (ncRNAs) are transcribed from active CREs. The function of these transcripts and their role in gene regulation is an area of active research[1,7-10]. We previously demonstrated that differential enhancer transcription can be used to define a gene regulatory network (GRN)[1]. We identified ncRNAs whose expression was dependent on the cardiac transcription factor TBX5[1]. These genome wide TBX5-dependent ncRNAs from the adult mouse atria defined a GRN for cardiac rhythm[1]. This approach led to the identification of potent regulatory elements in cardiomyocytes and identified functional ncRNAs that mediated TBX5-dependent gene regulation. These findings suggested that differential enhancer transcription may be an effective complementary approach to chromatin accessibility, epigenetic marks, and reporter assays for the discovery of context-dependent CREs. However, the applicability of this approach across broader

79 contexts and its ability to distinguish activated and repressed elements has not been
80 examined.

81 Here, we applied the differential ncRNA approach to identify regulatory elements
82 that are active in either of two specific cell types, rod and cone photoreceptors, in the
83 mouse retina. Rod photoreceptors are active in dim light and constitute the most
84 abundant retinal cell type, comprising ~80% of all mouse retinal cells and 95% of human
85 photoreceptors[11,12]. In contrast, cone photoreceptors are active in bright light and
86 mediate high-acuity vision and color vision. Their critical role in daylight vision makes
87 them a desirable cell type to replace from stem cells, or to target for gene therapy, in
88 diseases that lead to blindness[13]. An understanding of the GRNs that control cone
89 versus rod fate is essential for understanding both normal retinal biology and for cone
90 replacement, as has been recently demonstrated for rods in mice[14-19]. In addition, gene
91 therapy vectors that require expression specifically in rods and/or cones would benefit
92 from a broader range of validated photoreceptor CREs[20-22].

93 Rods and cones are produced by retinal progenitor cells (RPCs), with cones
94 generally produced earlier in development than rods, from RPCs that express the TFs
95 *Otx2*, *Olig2*, and *Oc1*[23-26]. These and other TFs with essential roles in RPCs, rods, or
96 cones have been identified[27,28]. In addition to *Otx2*, a close homologue, *Crx,* is required
97 for normal gene expression in both rods and cones[29-31]. In contrast, *Nrl*, a basic leucine
98 zipper TF, is expressed only in rods and is required for their formation. In *Nrl*$^{-/-}$ mutant
99 mice*,* rod photoreceptors fail to form and are instead transformed into cells that
100 resemble cone photoreceptors in most respects[32].

We aimed to identify the photoreceptor CREs that regulate rod and cone gene expression programs using differential ncRNA transcriptome analysis. We compared ncRNA abundance genome-wide in the wild-type versus *Nrl*$^{-/-}$ mouse retina. We found that *Nrl*-activated versus repressed ncRNA transcripts defined photoreceptor regulatory elements. *Nrl*-activated ncRNAs, predominant in the wild-type retina, identified locations of accessible chromatin in rods and that were near rod-expressed genes. In contrast, *Nrl*-repressed ncRNAs, predominant in the *Nrl*$^{-/-}$ retina, identified locations of accessible chromatin in cones that were near cone-expressed genes. Moreover, differential expression of *Nrl*-dependent ncRNAs and that of local target genes were quantitatively correlated. Furthermore, the change in expression of these ncRNAs positively correlated with differential signal from rod and cone ATAC-seq. Direct assessment of *Nrl*-repressed loci identified active elements for photoreceptor expression, enriched for cone-specific genes. These data illustrate the utility of differential ncRNA profiling for nominating TF-dependent and context-specific regulatory elements.

Results

Nrl-dependent coding and non-coding transcriptional profiling identifies photoreceptor CREs

We interrogated *Nrl*-dependent coding and non-coding transcription in the mouse retina. We performed mRNA transcriptional profiling to determine *Nrl*-dependent coding gene expression (Figure 1A). We sequenced cDNA libraries made from polyA+ selected RNA from retinas of litter-matched WT and *Nrl* mutant adult mice at postnatal day 21 (P21) ($Nrl^{+/+}$ vs $Nrl^{-/-}$, n=5 and n= 6 resp., Figure 1A). By P21, mouse photoreceptor differentiation is largely complete. Genotype described 98% of the variance between samples, indicating a specific effect mediated by *Nrl* deletion across biological replicates (Figure 1B). Differential expression testing revealed 4,315 misregulated genes (Figure 1A). *Nrl* expression was absent from $Nrl^{-/-}$ samples, along with numerous known *Nrl* targets and rod-specific genes, including *Rho*, *Gnat1*, and *Nr2e3*. Conversely, genes whose expression is normally absent in rods, including those involved in cone differentiation showed activation in $Nrl^{-/-}$ samples, including *Gnat2, Gngt2*, and *Opn1sw*[18]. These expression changes were consistent with previously identified *Nrl*-dependent gene expression and rod- versus cone-specific gene expression patterns in the adult retina[16-18,33,34].

We performed ncRNA transcriptional profiling to identify *Nrl*-dependent ncRNAs. We performed deep sequencing of non-polyadenylated RNA from the same control and *Nrl* mutant retinal samples described above (Figure 1C)[35]. This approach identified approximately 30,000 retinal non-coding transcripts by *de novo* transcript assembly (Supplemental FIgure). Approximately 4,657 of these transcripts were *Nrl*-dependent

intergenic ncRNAs (FDR<0.05, FC>2). Genotype described 98% of the variance between samples, indicating a specific effect mediated by *Nrl* deletion (Figure 1D). Of the 4,657 *NRL*-dependent ncRNAs, 3,112 were significantly downregulated, or *Nrl*-activated, and 1,545 were significantly upregulated, or *Nrl*-repressed (Figure 1C).

We hypothesized that some *Nrl*-dependent noncoding transcripts marked *Nrl*-dependent retinal enhancers. Active regulatory elements are characterized by characteristic genomic patterns, including open chromatin[3-5]. To examine the epigenomic landscape of mouse retinal tissue, we analyzed Assay for Transpose Accessible Chromatin (ATAC-Seq) data from whole mouse wild-type retinal tissue at P21 (GSE72550)[19,36,37]. *Nrl*-dependent ncRNAs significantly overlapped with locations of accessible chromatin in the wild-type retina (Figure 1E, top).

We attempted to affiliate *Nrl*-dependent ncRNA-defined regulatory elements with candidate target *Nrl*-dependent coding genes. We sought *Nrl*-dependent mRNAs within 2MB of the ncRNA locus, a conservative consideration of potential distance parameters for CREs[1]. We performed Gene Ontology (GO) analysis of the *Nrl*-dependent genes associated with *Nrl*-dependent ncRNAs. This analysis described enrichment for GO terms related to phototransduction, sensory perception, and response to light stimulus, all consistent with photoreceptor gene expression (Figure 1E, bottom).

We and others have previously hypothesized that quantitative changes in CRE transcription mirrors quantitative changes in CRE activity[1,38]. Identification of CREs using ncRNA transcriptional profiling affords the determination of context-dependent quantitative changes in CRE transcription. We examined the quantitative correlation between the change in *Nrl*-dependent ncRNA transcription and the change in the most

proximal *Nrl*-dependent gene. First, we observed that the directionality of the change of each *Nrl*-dependent ncRNA and its local mRNA were significantly concordant. Secondly, the relative expression of *Nrl*-dependent ncRNAs was positively correlated with the relative expression of the most proximal NRL-dependent gene (Figure 1F Cor=0.58, P=9.7e-74). Together, these observations indicate a positive quantitative relationship between enhancer transcription and target gene regulation. Together these observations indicate that *Nrl*-dependent ncRNA profiling identifies genomic regions overlapping with open chromatin and that quantitatively correlate with local *Nrl*-dependent gene expression.

Nrl-dependent ncRNA transcriptional profiling identifies two distinct gene regulatory signatures, specific to rod and cone photoreceptors

We hypothesized that removal of *Nrl* revealed two distinct regulatory networks: a pathway driven by *Nrl* in the *wild-type* retina, composed of genes that require *Nrl* directly or indirectly and are thereby downregulated in $Nrl^{-/-}$ samples, referred to as *Nrl*-activated genes; and a pathway that emerged in the *Nrl* mutant retina, composed of genes whose transcription is directly or indirectly negatively regulated by *Nrl* and are thereby upregulated in $Nrl^{-/-}$ samples, referred to as *Nrl*-repressed genes. NRL is a well-described driver of rod differentiation and repressor of cone differentiation [16,32,39-41]. We therefore hypothesized that *Nrl*-activated ncRNAs, higher in the wild-type retina, would specifically define CREs active in rods, whereas the *Nrl*-repressed ncRNAs, higher in the *Nrl* mutant retina, would specifically define CREs active in cones.

We attempted to identify candidate rod and cone CREs by comparing *Nrl*-dependent ncRNAs with open chromatin signatures from rod and cone specific cell populations (GSE83312)[18], as opposed to whole retinal tissue (Figure 1E top). We found that both *Nrl*-activated and *Nrl*-repressed ncRNAs overlap with open chromatin in both rods and cones (Figure 2A-B, top). However, the *Nrl*-activated ncRNAs showed much greater enrichment for open chromatin in rods relative to open chromatin in cones (Figure 2A-B, OR=2.77 vs 2.36). Conversely, *Nrl*-repressed ncRNAs showed much greater enrichment with cone open chromatin than with rod open chromatin (Figure 2A-B, OR = 5.15 vs 3.35). We interrogated the distribution of activated and repressed *Nrl*-dependent ncRNAs with chromatin accessibility to define rod-specific, cone-specific, and shared CREs. By overlapping rod and cone ATAC-seq sets, we identified total regions of accessible chromatin, of which 25,666 were shared between rods and cones, 9,116 were rod-specific, and 9,148 were cone-specific (Figure 2C). Intersection of *Nrl*-dependent ncRNAs with each group revealed that *Nrl*-activated and *Nrl*-repressed ncRNAs most frequently emanated from regions that were accessible in both rods and cones, suggesting that the majority of Nrl-dependent elements are shared photoreceptor elements common to all photoreceptor types (Figure 2C). However, when comparing the pattern of *Nrl*-dependent ncRNA expression from regions accessible in only rods or cones, a cell type-specific pattern emerged: *Nrl*-activated ncRNAs overlapped more frequently with regions only accessible in rods versus cones (105 vs 38, p=7.9e-10, Figure 2C), whereas *Nrl*-repressed ncRNAs overlapped more frequently with regions only accessible in cones versus rods (Figure 2C, 137 vs 10, $p < 2.2e-16$). These comparisons indicate that *Nrl*-dependent ncRNAs can be separated into three

distinct bins of *Nrl*-dependent candidate CREs; *Nrl*-activated or *Nrl*–repressed ncRNAs at shared rod and cone accessible regions; *Nrl*-activated ncRNAs, expressed in the wild-type retina, at regions accessible only in rods, and *Nrl*-repressed ncRNAs, expressed in the *Nrl*$^{-/-}$ retina, at regions accessible only in cones.

We asked if *Nrl*-activated and *Nrl*-repressed ncRNA expression levels were concordant with the predicted strength of rod vs. cone CREs, as defined by the local quantitative enrichment of ATAC-seq read density (Figure 2D). We compared *Nrl*-activated and *Nrl*-repressed ncRNAs to wild-type and *Nrl*-deficient ATAC-seq reads (Figure 2D, left). We observed that CREs with *Nrl*-activated ncRNAs were associated with significantly higher ATAC-seq read density in the WT retina compared with the *Nrl*$^{-/-}$ retina (Figure 2D; p=3.1e-5) (GSE72550)[19]. Conversely, we observed that CREs with *Nrl*-repressed ncRNAs were associated with significantly higher ATAC-seq read density in the *Nrl*$^{-/-}$ retina compared with the WT retina (Figure 2D, p= 4.5e-10). (GSE72550)[19].

We next assessed the association of *Nrl*-activated and *Nrl*-repressed ncRNAs with ATAC-seq read density in sorted rods and cones (GSE83312). We found that *Nrl*-activated ncRNAs were enriched at regions of higher ATAC-seq signal in rods than in cones (Figure 2D; p= 2.3e-5). In contrast, *Nrl*-repressed ncRNAs were enriched at regions of higher ATAC-seq signal in cones than in rods (Figure 2D; p= 5.3e-11). These observations indicate that *Nrl*-dependent ncRNAs differentially associate with candidate rod versus cone CREs: *Nrl*-activated ncRNAs affiliated with strong rod ATAC-seq signal, whereas *Nrl*-repressed ncRNAs affiliated with strong cone ATAC signal.

We hypothesized that the *Nrl*-dependent change in ncRNA may correlate with cell type specific rod versus cone change in chromatin accessibility. We therefore

231 compared the *Nrl*-dependent change in ncRNA transcription at rod- and cone-specific
232 ATAC-seq regions to the relative ATAC-seq read enrichment in rods vs. cones at those
233 locations. We observed a positive correlation between downregulated ncRNAs and rod
234 ATAC-seq reads, and between upregulated ncRNAs and cone ATAC-seq reads (Figure
235 2E, Cor=0.6, P<2.2e-16). This result indicated that the direction and quantitative
236 degree of *Nrl*-dependence of ncRNA expression correlated with the relative cell-type
237 specificity of the ATAC-seq signal, for *Nrl*-activated ncRNAs with rod-specific ATAC-seq
238 and *Nrl*-repressed ncRNAs with cone-specific ATAC-seq.

239 We compared regions with high-density putative regulatory elements[43,44] from
240 rod and cone ATAC-seq with the *Nrl*-activated and *Nrl*-repressed ncRNAs. We
241 observed that *Nrl*-activated ncRNAs predicted clusters of putative rod enhancers
242 (Figure 2F, OR = 1.7, p = 4.7e-7) but not cone (Figure 2F, OR 1.05, p = 0.64). In
243 contrast, *Nrl*-repressed ncRNAs predicted high density regulatory elements[43,44] from
244 cone (Figure 2F, OR = 1.6, p = 1.9e-4) but not rod cells (Figure 2F, OR = 0.95, p =
245 0.87). We conclude that *Nrl*-activated and repressed ncRNAs enrich two distinct
246 regulatory pathways, identifying strong candidate CREs specific to rod and cone
247 photoreceptors, respectively.

248 Mouse photoreceptor CREs are highly enriched for binding sites of CRX, a
249 homeobox TF required for both rod and cone gene expression[14,17,42]. We asked if CREs
250 defined by *Nrl*-dependent ncRNAs were CRX-bound. We observed significant
251 enrichment for CRX occupancy by ChIP-seq in wild-type whole retinal tissue at *Nrl*-
252 dependent ncRNA-defined regions of accessible chromatin (Figure 3A, top,
253 GSE20012)[15]. Interestingly, using ATAC-seq data in conjunction with ncRNA

transcription did not improve identification of CRX-bound elements (Figure 3A, bottom), suggesting that differential ncRNA transcription independently identifies CREs likely to be bound by CRX, without the need for CRX ChIP. Regions characterized as either *Nrl*-activated or repressed from the ncRNA analysis of the *Nrl*$^{-/-}$ retina identified genomic CRX localization in wild-type tissue (Figure 3B, left). In *Nrl*$^{-/-}$ retina, only *Nrl*-repressed ncRNAs identified CRX localization, providing candidates for cone-specific CRX occupancy, as expected with the absence of rod fate in the *Nrl*$^{-/-}$ retina, and consistent with analysis of CRX occupancy in *Nrl*$^{-/-}$ retinas (Figure 3B, right)[15].

We performed *de novo* motif analysis of *Nrl*-dependent ncRNA-defined CREs to identify potential transcriptional co-regulators. Both activated and repressed ncRNAs identified the K50 homeodomain motif (K50 indicating the presence of lysine at position 50 of the homeodomain), shared among known NRL co-regulators, CRX and OTX2[15,18,45] (Figure 3C, row 1). We interrogated the motifs specific for *Nrl*-activated and *Nrl*-repressed CREs, attempting to identify motifs specific for rod or cone CREs, respectively. *Nrl*-activated CREs were significantly enriched for the NRL binding motif (bZip) and NR2E3 (NR), a direct downstream target of NRL[37] (Figure 3C, rows 2 and 4). By contrast, *Nrl*-repressed ncRNAs were differentially enriched for NeuroD1 binding motif (BHLH, Figure 3C, row 3). Thus, *Nrl*-dependent ncRNAs identify *Nrl* activated and repressed CREs characterized by specific TF binding motifs.

To define potential novel transcriptional pathways that are unique to rod and cone photoreceptor cell types, we assessed comparative motif enrichment for *Nrl*-activated CREs in rod-specific open chromatin versus *Nrl*-repressed CREs in cone-specific open chromatin (Figure 3D-E). We identified TF motifs unique to rods (Figure

3D) and cones (Figure 3E). Consistent with comparative analysis in Figure 3C, we find the NR motif unique to rod-specific elements (Figure 3D), and the bHLH binding motif unique to cone-specific regulatory elements (Figure 3E) with this comparative analysis. These differential motifs predict rod and cone specific TFs that may help define the two cell types.

Nrl-dependent ncRNAs define functional photoreceptor regulatory elements

We predicted that the ncRNA-defined candidate regulatory elements are active in photoreceptors, with at least a subset differentially active in rods versus cones. We first examined the regulatory capacity of *Nrl*-dependent ncRNA-defined CREs in the developing mouse retina. As *Otx2* is required for the genesis of rods and cones, the regulation of this locus is of great interest[25]. Data from the ENCODE project allowed us to identify 24 DNase I hypersensitivity sites (HS) at the *Otx2* locus, from mouse retina at P0, within 300kb around the *Otx2* gene (Figure 4A)[46]. These regions were compared with those identified through *Nrl*-dependent ncRNA profiling. Of the 24 ENCODE sites, 4 overlapped *Nrl*-repressed ncRNAs and cone ATAC-seq peaks (Figure 4B). We examined the activity of all 24 locations to determine if candidate CREs overlapping with *Nrl*-repressed ncRNAs would be distinguished from those without overlapping ncRNAs. The DNA sequences corresponding to the DNAse I HS sites were cloned into the reporter plasmid Stagia3, which has an eGFP-IRES-AP reporter[47,48]. We tested these constructs, along with a control, ubiquitously expressed CAG-Cherry plasmid, for activity in dissected mouse retina through electroporation at E14.5. E14.5 is a period of development when primarily cones, and not rods, are being generated, and

electroporation into mature cones at later ages is inefficient[49]. Retinas were then cultured as explants on filters for 2 days. Nine out of the 24 DNAse I HS sites showed alkaline phosphatase (AP) activity (Figure 4A, C). Interestingly, three of the four ENCODE DNAse I HS sites corresponding with *Nrl*-dependent ncRNAs with P21 cone ATAC-seq peaks were among the most active, highlighting the observation that TF-dependent ncRNAs mark potent regulatory elements (Figure 4B).

As cone development is regulated by both *Otx2* and Onecut 1 (*Oc1*), we also were interested in regulatory elements at the *Oc1* locus[25,50]. To this end, several regions predicted by the ENCODE DNAse I analysis were tested (Figure 4D). As with *Otx2*, a region with strong regulatory activity had an *Nrl*-dependent ncRNA. These results indicate that intersecting TF-dependent ncRNA expression with previously published ENCODE datasets provides novel information useful for identifying strong regulatory elements, compared to utilizing chromatin accessibility alone.

We hypothesized that *Nrl*-repressed ncRNAs may identify CREs with cone activity. To examine whether CREs defined by the overlap of *Nrl*-repressed ncRNAs and cone ATAC-seq peaks have activity in developing cones, we examined ncRNA defined elements from the *Otx2* and *Rxrg* loci, as these genes are known to be important for cone development[30,51]. The DNA sequences from these loci (material and methods) were cloned into the Stagia3 reporter plasmid. We also tested putative CREs defined by ncRNA expression for other genes involved in the development of the visual system (*En2, Socs3, Nab1, Six6, Opn1sw*), as well as *Pde6b*, a rod-specific gene. These plasmids were delivered by electroporation of ex-vivo E14.5 mouse retinas and assayed for AP activity 2 days later (Figure 4E). The ThrbCRM1-dtTomato construct

was used as a positive control for AP, as it has known activity in cones, HCs, and a subset of RPCs that produce cones and HCs in the chick retina[25]. Ten of 12 candidate regulatory elements were able to drive AP expression in explanted E14.5 mouse retinal tissue (Figure 4E).

To define the specific cell-types with reporter activity, active enhancer constructs were tested for expression of eGFP, which provided greater cellular resolution and coincident assay with cell-type specific marker co-expression. The ThrbCRM1-tdTomato plasmid was first tested for its use as a positive control to specifically mark murine cones. ThrbCRM1-tdTomato positive cells were located in the apical region of electroporated E15.5 retinas (Figure 5A), which is the location of developing cones[18,25], and they were positive for the Rxrg protein, a validated cone marker (Figure 5B). We then examined co-expression of eGFP, driven by the *Nrl*-repressed ncRNA constructs, and tdTomato from ThrbCRM1 following electroporation into E14.5 retinal explants. GFP-positive cells from 7 of the *Nrl*-repressed ncRNA constructs were located in the apical region and showed a strong overlap with tdTomato (Figure 5C). The only exception was the *Socs3* element (Figure 5C), which marked cells with a morphology and position matching those of RPCs. We further tested the expression from the *Otx2* ncRNA-defined enhancer regions 1 and 2 with that of the OTX2 protein (Figure 5D). Both *Otx2*-ncRNA-defined elements showed strong co-localization with the Otx2 protein. Together, these findings indicate that CREs marked by *Nrl*-dependent ncRNAs in the *Nrl* mutant setting have activity in developing cones.

Discussion

Application of enhancer transcription to the definition of *Cis*-Regulatory Elements

Defining the GRNs that distinguish healthy and disease states requires the identification of the functionally relevant CREs and TFs that regulate them. Current approaches for nominating CREs, including histone modifications, chromatin status[3-5] and TF occupancy[6], have been successful in many contexts. However, only a small fraction of the thousands of candidate TF-dependent enhancers identified by these approaches have been functionally validated[52]. Moreover, these approaches do not reveal context specificity of expression from a region, nor provide a quantitative assessment of enhancer function[1,53]. We previously defined a complementary approach to regulatory region identification, utilizing context-dependent enhancer transcriptional profiling, to nominate regulatory elements. This approach defined a GRN composed of functional elements in cardiomyocytes defined by ncRNA expression from the adult mouse heart that contribute to cardiac rhythm control[1]. Here we extended TF-dependent ncRNA profiling to define regulatory regions that govern photoreceptor gene expression in the retina and assess the applicability of this approach for defining both wild-type and mutant elements that comprise context-specific GRNs.

Regulatory regions active in photoreceptors and their progenitor cells

An understanding of the GRNs that govern photoreceptor production is particularly important, given the desire to generate cones from stem cells for therapeutic applications[13]. Definition of the GRN for cone genesis has been limited, as cones are typically born early, when access is limited, and cones are much less abundant than

rods. The conversion of rods to cone-like cells in the *Nrl*$^{-/-}$ mouse has provided a deeper understanding of the molecular components of both cones and rods, including the identification of rod- and cone-enriched transcripts[18,28], as well as the delineation of rod and cone open chromatin regions[18,19]. These datasets have provided an excellent background for an assessment of TF-dependent ncRNA profiling as a method for nominating regulatory regions near *Nrl*-dependent genes.. Consistent with a correlation between ncRNA transcription and regulatory region activity, the *Nrl*-dependent ncRNAs defined regions that had a high density of marks of regulatory regions, such as those proximal to *Rxrg, Otx2*, and *Gas6*[7-10, 43,44].

Nrl-dependent ncRNAs that were upregulated in the *Nrl* mutant retina identified candidate regulatory regions associated with cone genes, and ncRNAs that were downregulated in the *Nrl* mutant identified candidate regulatory regions associated with rod genes. Functional examination of enhancer activity for a subset of these predicted elements at relevant cone genes showed cell type specificity for the majority of *Nrl*-repressed elements (Figure 5). However, a small number of tested elements did not display the predicted cone pattern of activity. A few had activity in other cell types, or had no specific activity in retinal explants. Some DNase I hypersensitivity sites from the Otx2 locus where we found *Nrl*-repressed elements also had activity in bipolar interneurons (not shown, Wang et al. in preparation), which are born in the postnatal period, and in rods[54]. This may be due to the fact that the CREs defined by ncRNAs and ATAC-seq peaks are typically fairly large and thus may harbor multiple TF binding sites. These binding sites may be rod, cone, or bipolar cell-specific, and may rely on higher ordered chromatin structure for proper regulation, structure that likely is not included in

392 electroporated plasmids. Elements that showed no specific activity in retinal explants
393 may be active in mature cones, and not in the cone progenitor cells or immature cones
394 that were assayed here.

395 While many of the genes associated with the ncRNA-repressed list were related
396 to cone development or function (*Otx2, Rxrg, Gngt2, Gnat2, Opn1sw, Sall3*)[30,51,55-57], we
397 also found genes that are generally important for eye or retina development that have
398 not been well characterized with respect to cone-specific development. Interestingly, the
399 ncRNA-defined CRE associated with *Six6*, an eye-field TF, or the element near *En2*, a
400 gene important for ganglion cell differentiation[18,58], displayed reporter activity in cones.
401 These results suggest that non-coding transcriptional profiling uncovered not only rod-
402 and cone-specific regulatory programs, but potential shared regulatory programs that
403 warrant further investigation.

404

405 **Defining TF-dependent networks of cell fate has implications for the study of**
406 **disease-specific regulatory pathways**

407 Defining both activated and repressed regulatory programs has significance
408 across biological contexts for distinguishing the CREs driving gene regulation in normal
409 versus mutant, or disease, states. Disease GRNs have been primarily characterized via
410 discovery of wild-type enhancers, followed by assessment of disease-specific changes
411 in their activity. This presumes a model in which disruption of the wild-type GRN
412 sufficiently describes the disease state. However, our findings highlight an emergent,
413 mutant-specific, GRN through the use of context-specific CRE activity. Defining
414 regulatory pathways in NRL presence and absence parallels potential studies of healthy

415 versus diseased states. Defining such context-specific CREs in disease states may
416 define novel regulatory pathways that mediate disease pathogenesis, with potential
417 therapeutic implications. The effectiveness of enhancer transcriptional profiling for
418 identifying emergent networks suggests a potential for future application, for the
419 assessment of both cell-type specific and disease GRNs.
420
421

422 **Acknowledgments:** This research was supported in part by NIH R01 HL092153,
423 R01HL124836, and R33 HL123857 to IPM, R01EY024958, R01EY025196, and
424 R01EY026672 to JCC, F30 HL131298 to RDN, AHA Collaborative Sciences Award to
425 IPM, and HHMI (CLC). This research was supported in part by the Leducq Foundation
426 (IPM). This research was supported in part by the NIH through resources provided by
427 the Computation Institute and the Biological Sciences Division of the University of
428 Chicago and Argonne National Laboratory, under grant 1S10OD018495-01. N.L. was
429 supported by post-doctoral fellowships from the Swiss National Science Foundation and
430 the Human Frontiers Science Program.
431

Materials and Methods

Animals

Nrl$^{-/-}$ mice were generated as previously described[18,59]. Mouse husbandry and all procedures (including euthanasia by CO_2 inhalation and cervical dislocation) were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, and were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee. For ex-vivo enhancer testing, wild-type embryos were obtained from timed pregnant CD1 mice (Charles River Laboratories). All animal studies were approved by the Institutional Animal Care and Use Committee at Harvard University.

Coding RNA-Seq library preparation and Data analysis

Libraries were prepared from this RNA starting with 1 g per sample and using the mRNA-seq Sample Prep Kit (Illumina) as per recommended instructions. After Ribozero purification and removing only ribosomal RNA, barcoded libraries were prepared according to Illumina's instructions (2013) accompanying the TruSeq RNA Sample prep kit v2 (Part# RS-122-2001). Libraries were quantitated using the Agilent Bio-analyzer (model 2100) and pooled in equimolar amounts. The pooled libraries were sequenced with stranded 50-bp single-end reads on the HiSeq2500 in Rapid Run Mode following the manufacturer's protocols (2013).

RNA library preparation was performed as previously discussed[1]. Briefly, 22M to 30M reads were mapped to mouse genome with TopHat2 (v 2.1.1). Reads mapped to the mitochondrial genome, and with phred score < 30 were excluded. Counts were retrieved with HTseq (v.0.6.0)[60] in union mode. Lastly, counts were analyzed for differential expression with R (3.4) package DEseq2[61].

Noncoding RNA-Seq library preparation

Total RNA was extracted by TRIzol Reagent (Invitrogen), followed by ribosomal and polyA depletion. After RiboZero purification and oligo-dT depletion, RNA Barcoded Libraries were prepared according to Illumina's instructions (2013) accompanying the TruSeQ RNA Sample prep kit v2 (Part# RS-122-2001). Libraries were quantitated using the Agilent Bio-analyzer (model 2100) and pooled in equimolar amounts. The pooled libraries were sequenced with 50-bp stranded single-end reads on the HiSEQ4000 in Rapid Run Mode following the manufacturer's protocols (2013).

Noncoding RNA-Seq Data analysis

About 170–186 million high-quality reads (quality score >30) reads for each sample were obtained. Fastq files were aligned to UCSC genome build mm9 using TopHat (version 2.0.10) as previously described[62] and between 168 million and 174 million reads were successfully mapped. RABT assembly was performed by Cufflinks (version 2.2.1, with parameters –g --frag-bias-correct --multi-read-correct --upper-quartile-norm), as it can recover transcripts that are transcribed from segments of the genome that are missing from the current genome assembly. Analysis of differential expression was performed using Cuffdiff from the Cufflinks package[62]. False discovery rate (FDR) was

476 calculated after removing the coding-gene counts. Significance was considered to have
477 been reached when FDR was <0.05 and fold change was > 2. The mm9 genomic
478 coordinates of identified noncoding transcripts were lifted over to the mouse mm10
479 before comparisons with open chromatin and TF-binding regions.

480 *GO enrichment analysis*

481 Enrichment of GO Biological Process terms from genes within 2MB of ncRNAs was
482 performed with Bioconductor package GOstats version 2.46[63].

483 *ATAC-seq and ChIP-seq data processing*

484 Fastq files from previously generated ChIP-seq (GSE20012) and ATAC-seq
485 (GSE83312) datasets were downloaded from GEO and processed identically as
486 previously described[18]. Briefly, adapter sequences were clipped from reads using
487 Cutadapt[64], then aligned to UCSC mouse genome mm10 with bowtie version 2.3.4[65] in
488 end-to-end mode. Mismatched reads, PCR duplicates, ENCODE blacklisted regions,
489 and reads with quality < 30 were removed with Samtools version 1.5[66]. For ATAC-seq,
490 fragments with width > 147 base pairs were removed to enrich for nucleosome free
491 reads using a custom script. Peaks for both assays were called with Macs 2.11[67].

492 *Associating ncRNAs and regulatory elements*

493 Open chromatin peaks and TF-binding peaks were intersected with ncRNAs using
494 Bioconductor package GenomicRanges[54] allowing for a 500bp gap.

Metagene analysis

To compare the coverage of ChIP-seq and ATAC-seq regions in differentially expressed ncRNAs we used Bioconductor Package Metagene (2.12.1). Coverage was normalized to reads per million. We binned the position of each region to 100 bp. We modified the current Metagene source code to output boxplots as opposed to ribbons and we then tested the difference of means with ANOVA.

Identification of differentially accessible peaks

Rod and cone ATAC-seq peak regions were combined and sorted with bash commands (cat sort). Counts were retrieved from each alignment file using Bedtools multicov (2.26.0) and tested for differential expression with DESeq2^. Peaks were considered differentially accessible when log2 fold change greater than 1 and p-adjusted value less than 0.05 and these regions were considered cone-specific and rod-specific. To assess the relationship between differentially expressed ncRNAs and differentially accessible open chromatin we performed a global overlap of all ncRNA regions and combined ATAC-seq regions. Then differential regions from both sets were highlighted.

Motif Enrichment Analysis

Known and de novo motif scanning was performed with HOMER (4.3) using Rod-specific and cone-specific ATAC-seq regions that intersected with *Nrl* wildtype ncRNAs and *Nrl* null ncRNAs respectively. Target sequences consisted of 200 bp elements centered on peak summits. Background sequences consisted of approximately 50,000 randomly selected 200 bp intervals from the mouse genome normalized for mono- and di-nucleotide content relative to each target set. Repeat sequences were masked from

518 the genome, and targets with >70% of bases masked were dropped from enrichment
519 analysis. Preferential spacing between highly enriched motifs (MAF, bZIP, and NR
520 motifs) in differential regions was assessed by first centering the above intersect on
521 individual motifs and plotting the density.
522
523 *Identification of high density peaks regions*
524 Cone and Rod specific high density peak regions were marked by HOMER (4.3) using a
525 12.5kb window[43,44]. Alignment files for rod and cone ATAC-seq were used to find peaks
526 with Homer using the parameter –style Super.
527
528 *Retina electroporation and AP staining*
529 Ex vivo retina electroporation was carried out as described previously[68,69], with at least
530 three biological replicates for AP staining, or at least in duplicates for
531 immunohistochemistry. The chamber used for electroporation was modified as
532 previously described[70]. Stages of embryos used for the experiments are described in
533 the main text of in the figure legends. Electroporation settings were 5x30 V pulses, 50
534 ms each and 950 ms apart. DNA concentration was 400-600ng/ul for control plasmids
535 and 1ug/ul for enhancer constructs. Retinas were harvested after 2 days in culture.
536
537 *Plasmid and DNA sequences*
538 In vivo enhancer testing was performed with the Stagia3 reporter vector (Addgene
539 #28177)[48]. Enhancer testing with the CAG-EGFP, CAG-mCherry, and ThrbCRM1-

tdTomato vectors were modified from our previous work[25,69]. Coordinated of regions cloned are shown in mm10 assembly.

O10> chr14:48616314-48617497

O11> chr14: 48624486-48626389

O5> chr14: 48579937-48581029

O6> chr14: 48584564-48585012

O7> chr14: 48593170-48594188

O8> chr14: 48606973-48608016

O9> chr14: 48608144-48609697

O15> chr14:48662841-48663211

O20> chr14:48740203-48742409

Oc1 A> chr9:74384085-74384740

Oc1 B> chr9 74530189-74532399

Oc1 C> chr9:74810971-74812406

Otx2 ncRNA1>chr14 + 48580418 48580655

Otx2 ncRNA2>chr14 + 48593310 48594038

Otx2 ncRNA3>chr14 + 48608844 48609310

Otx2 ncRNA4>chr14 + 48617045 48617321

Rxrg ncRNA>chr1 + 167438156 167438645

Pde6b ncRNA>chr5 + 108366779 108367075

En2 ncRNA>chr5 + 28145373 28145644

Socs3 ncRNA>chr11 + 117963224 117963787

Nab1 ncRNA>chr1 + 52435750 52436194

Six6 ncRNA1>chr12 + 72854405 72854809

Six6 ncRNA2>chr12 + 72831804 72832279

Opn1sw ncRNA>chr6 + 29394311 29394823

IHC

20-30um retinal section were prepared and stained as described previously[68]. Blocking solution was: 0.3% Triton in 1× PBS. Primary antibodies used in this study include: chicken anti-GFP (1:1000, Abcam, AB13970), rabbit anti-mCherry (1:1000, Abcam, 167453), rabbit anti-Otx2 (1:500, Millipore, AB9566), mouse anti-Rxrg (1:300, Santa Cruz Biotechnology, sc-514134). Secondary antibodies were from Jackson Immunoresearch.

Imaging

Retina explants were imaged on a Leica M165FC microscope. Retinal section images were acquired by a Zeiss LSM780 inverted confocal microscope from the Microscopy Resources on the North Quad (MicRoN) core at Harvard Medical School.

Figure legends

Figure 1: Transcriptional profiling identifies *Nrl*-dependent coding and non-coding RNAs

A) Volcano-plot (log_2 (fold-change) vs $-log_{10}$ (P_{Adj})) of $Nrl^{+/+}$ vs $Nrl^{-/-}$ coding transcripts. Significantly misregulated genes ($P_{Adj} < 0.05$) in blue, and non-significant in grey. Selected significant coding genes known to play a role in photoreceptor differentiation labeled (black).

B) Principle component analysis on WT and $Nrl^{-/-}$ coding transcripts. Samples segregate primarily by genotype, describing 98% of variance on PC1.

C) Volcano-plot (log_2 (fold-change) vs $-log_{10}$ (P_{Adj})) of $Nrl^{+/+}$ vs $Nrl^{-/-}$ non-coding RNAs (ncRNAs). Significantly misregulated ncRNAs ($P_{Adj} < 0.05$) are depicted in blue and non-significant in grey. Selected transcripts labeled by nearest *Nrl*-dependent coding genes (black).

D) Principle component analysis on WT and $Nrl^{-/-}$ noncoding transcripts. Samples segregate primarily by genotype, describing 98% of variance on PC1.

D) Top: Venn diagram of the overlap of ATAC-seq from control mouse retina (GSE72550) and *Nrl*-dependent ncRNA transcripts (OR = 5.56, P < 2.2e-16). P-values and Odds Ratios from Fisher Exact Test. Bottom: Gene ontology (GO) analysis of nearby (2Mb) *Nrl*-dependent genes to defined *Nrl*-dependent ncRNAs. Fisher Odds (gray) and –log(FDR) (Red). Significantly enriched GO terms are related to visual processes, including “rhodopsin mediated signaling”, “phototransduction”, and “detection of light stimulus”.

603 F) Scatterplot in hexagonal binning for the *Nrl*-dependent ncRNAs. Differential
604 expression fold change (log2) of nearest *Nrl*-dependent gene vs differential expression
605 fold change (log2) of ncRNAs (x-axis).
606
607 **Figure 2: *Nrl*-activated and repressed ncRNAs identify cell-type specific**
608 **regulatory programs**
609 A) Venn diagrams of the overlap of *Nrl*-activated (top) and *Nrl*-repressed (bottom)
610 ncRNAs with ATAC-seq from rod photoreceptors (GSE83312). P-values and Odds
611 Ratios from Fisher Exact Test.
612 B) Venn diagrams of the overlap of *Nrl*-activated (top) and *Nrl*-repressed (bottom)
613 ncRNAs with ATAC-seq from sCone photoreceptors (GSE83312). P-values and Odds
614 Ratios from Fisher Exact Test.
615 C) Venn diagram depicting the intersections of ATAC-seq from rods and ATAC-seq
616 from sCones (GSE83312) with *Nrl*-repressed and *Nrl*-activated ncRNAs. P-values and
617 Odds Ratios from Fisher Exact Test.
618 D) Boxplots depicting the enrichment of Mean Reads per Million (RPM) of ATAC-seq
619 from WT P21 mouse retinas and in *Nrl-/-* ATAC-seq P21 mouse retinas (left,
620 GSE72550) and of Rod-specific ATAC-Seq and sCone specific ATAC-Seq (right,
621 GSE83312) in *Nrl*-activated (green) and *Nrl*-repressed (blue) compared with
622 background (Bkg) non-*Nrl* dependent ncRNAs (red). P-value calculated by ANOVA
623 testing.
624 E) Scatterplot depicting correlation of differentially expressed ncRNA transcripts with
625 RPM of differentially expressed ATAC regions from rod (red) and cone photoreceptors

(blue). Log2 (sCone ATAC RPM/Rod ATAC RPM) vs log2(*Nrl*$^{-/-}$ ncRNA RPM/*Nrl*$^{+/+}$ ncRNA RPM). Non-significantly changed transcripts in gray.

F) Venn diagram depicting the overlap of annotated rod and cone specific elements as defined by high density reads[43,44,55] with *Nrl*-activated ncRNAs (left) and *Nrl*-repressed ncRNAs (right). P-values and Odds Ratios from Fisher Exact Test.

Figure 3: ncRNA-defined cell-type specific regulatory elements are enriched for cell-type specific TF motifs

A) Venn diagram of the overlap of total *Nrl*-dependent ncRNAs (top) and the intersection of Nrl-dependent ncRNAs with Retinal ATAC (GSE74661, bottom) with CRX bound sites from ChIPseq in whole mouse retinas (GSE20012).

B) Boxplots depicting the relative enrichment of *Nrl*-activated (green) and *Nrl*-repressed (blue) ncRNAs in control CRX ChIP retinas (left) and *Nrl-/-* CRX ChIP retinas (right), in relation to the ncRNA background (red). P value calculated by ANOVA testing.

C) Position Weight Matrix (PWM, top) of select *de novo* binding motifs identified by *Nrl*-dependent ncRNAs. Motif Density (bottom) as a function of distance from center (bp) in *Nrl*-activated (Blue) and *Nrl*-repressed (Red). Hbox motif (CRX, OTX), BHLH motif (NeuroD1, NeuroG2), NR (NR2E3) and bZip (NRL, Mafa) motifs shown.

D) Position Weight Matrix (PWM) of rod-specific *de novo* binding motifs grouped by TF family and ranked by p-value.

E) Position Weight Matrix (PWM) of cone-specific *de novo* binding motifs grouped by TF family and ranked by p-value.

Figure 4: *Nrl*-repressed ncRNAs identify cone-specific regulatory elements with regulatory activity *ex vivo*

A) Annotation track shows ENCODE DNaseI Hypersensitivity Sites for mouse retinal cells at various stages of development for the Otx2 locus. 24 peaks were selected for further testing (O1-O23) within a ~300kb region centered around Otx2.

B) Nrl-dependent ncRNAs at the Otx2 locus and the nearest ATAC-seq from cone P21 cells.

C) DNA sequences corresponding to the DNAse I HS peaks annotated in (A) were cloned into the reporter plasmid Stagia3 (alkaline phosphatase (AP) and EGFP) and tested for activity in the mouse retina by ex vivo electroporation at E14.5 along with a co-electroporation control CAG-Cherry. Retinas were then cultured as explant on filters for 2 days. Positive regions are shown in green in (A), while the negative are in red.

D) Electroporated constructs containing DNA sequences defined by retina DNAse I HS from ENCODE at the Onecut1 locus.

E) Retinal explants electroportated with ncRNA DNA sequences defined by cone ATAC-seq peaks associated with the ncRNAs found at the Otx2 locus and other known retinal genes.

Figure 5: *Nrl*-repressed ncRNAs identify functional cone-specific regulatory elements *in vivo*.

A) Transversal sections of retina from mice electroporated at E15.5 (A) or E14.5 (B-D) and cultured for 2 days before fixation and embedding are shown. The ThrbCRM1-dtTomato enhancer was electroporated into mouse retina with an ubiquitous CAG-GFP

672 co-electroporation control. ThrbCRM1 positive cells are located in the apical region of
673 the developing ONL, where photoreceptors are found (left). The co-electroporation GFP
674 reporter is expressed in mitotic cells at the time of the electroporation.
675 B) The ThrbCRM1 enhancer (left) was expressed in cells positive for Rxrg protein
676 (center), a cone marker gene.
677 C) Positive regions from our AP screen were co-electroporated (Figure 5 E) along
678 with the ThrbCRM1-dtTomato enhancer (left). The Stagia3 reporter plasmid used for AP
679 also contains an eGFP readout (center column). All but one of the cells positive for the
680 putative enhancers tested are located in the developing ONL, and show a strong
681 overlap with the cone marker ThrbCRM1.
682 D) Two Nrl-dependent ncRNAs at the Otx2 locus were tested (left column) for their
683 co-localization with the Otx2 protein (center). Both showed a strong overlap (composite
684 image, right column).
685
686
687

References

1 Yang, X. H. *et al.* Transcription-factor-dependent enhancer transcription defines a gene regulatory network for cardiac rhythm. *eLife* **6**, doi:10.7554/eLife.31683 (2017).

2 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

3 Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**, 148-156, doi:10.1038/ng.1064 (2012).

4 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

5 Werner, M. S. & Ruthenburg, A. J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell reports* **12**, 1089-1098, doi:10.1016/j.celrep.2015.07.033 (2015).

6 Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858, doi:10.1038/nature07730 (2009).

7 Davidovich, C. & Cech, T. R. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *Rna* **21**, 2007-2022, doi:10.1261/rna.053918.115 (2015).

8 Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455, doi:10.1038/nature20149 (2016).

9 Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-394, doi:10.1038/nature10006 (2011).

10 Wu, H. *et al.* Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLoS genetics* **10**, e1004610, doi:10.1371/journal.pgen.1004610 (2014).

11 Young, R. W. Cell differentiation in the retina of the mouse. *The Anatomical record* **212**, 199-205, doi:10.1002/ar.1092120215 (1985).

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

13 Oswald, J. & Baranov, P. Regenerative medicine in the retina: from stem cells to cell replacement therapy. *Ther Adv Ophthalmol* **10**, 2515841418774433, doi:10.1177/2515841418774433 (2018).

14 Aldiri, I. *et al.* The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis. *Neuron* **94**, 550-568 e510, doi:10.1016/j.neuron.2017.04.022 (2017).

15 Corbo, J. C. *et al.* CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. *Genome Res* **20**, 1512-1525, doi:10.1101/gr.109405.110 (2010).

16 Corbo, J. C., Myers, C. A., Lawrence, K. A., Jadhav, A. P. & Cepko, C. L. A typology of photoreceptor gene expression patterns in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12069-12074, doi:10.1073/pnas.0705465104 (2007).

17 Hsiau, T. H. *et al.* The cis-regulatory logic of the mammalian photoreceptor transcriptional network. *PLoS One* **2**, e643, doi:10.1371/journal.pone.0000643 (2007).

18 Hughes, A. E., Enright, J. M., Myers, C. A., Shen, S. Q. & Corbo, J. C. Cell Type-Specific Epigenomic Analysis Reveals a Uniquely Closed Chromatin Architecture in Mouse Rod Photoreceptors. *Scientific reports* **7**, 43184, doi:10.1038/srep43184 (2017).

19 Mo, A. *et al.* Epigenomic landscapes of retinal rods and cones. *eLife* **5**, e11613, doi:10.7554/eLife.11613 (2016).

20 Sinha, D., Phillips, J., Joseph Phillips, M. & Gamm, D. M. Mimicking Retinal Development and Disease With Human Pluripotent Stem Cells. *Investigative ophthalmology & visual science* **57**, ORSFf1-9, doi:10.1167/iovs.15-18160 (2016).

21 Aguirre, G. D. Concepts and Strategies in Retinal Gene Therapy. *Investigative ophthalmology & visual science* **58**, 5399-5411, doi:10.1167/iovs.17-22978 (2017).

22 Juettner, J. *et al.* Targeting neuronal and glial cell types with synthetic promoter AAVs in mice, non-human primates, and humans. *bioRxiv*, doi:10.1101/434720 (2018).

23 Altshuler, D. & Lillien, L. Control of photoreceptor development. *Current opinion in neurobiology* **2**, 16-22 (1992).

24 Gonzalez-Cordero, A. *et al.* Recapitulation of Human Retinal Development from Human Pluripotent Stem Cells Generates Transplantable Populations of Cone

Photoreceptors. *Stem Cell Reports* **9**, 820-837, doi:10.1016/j.stemcr.2017.07.022 (2017).

25 Emerson, M. M., Surzenko, N., Goetz, J. J., Trimarchi, J. & Cepko, C. L. Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors. *Developmental cell* **26**, 59-72, doi:10.1016/j.devcel.2013.06.005 (2013).

26 Hafler, B. P. *et al.* Transcription factor Olig2 defines subpopulations of retinal progenitor cells biased toward specific cell fates. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7882-7887, doi:10.1073/pnas.1203138109 (2012).

27 Corso-Diaz, X., Jaeger, C., Chaitankar, V. & Swaroop, A. Epigenetic control of gene regulation during development and disease: A view from the retina. *Progress in retinal and eye research* **65**, 1-27, doi:10.1016/j.preteyeres.2018.03.002 (2018).

28 Swaroop, A., Kim, D. & Forrest, D. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat Rev Neurosci* **11**, 563-576, doi:10.1038/nrn2880 (2010).

29 Inoue, T., Nishida, A. & Furukawa, T. [Transcriptional regulation of retinal photoreceptor cell development]. *Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme* **49**, 1413-1420 (2004).

30 Nishida, A. *et al.* Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. *Nature neuroscience* **6**, 1255-1263, doi:10.1038/nn1155 (2003).

31 Koike, C. *et al.* Functional roles of Otx2 transcription factor in postnatal mouse retinal development. *Molecular and cellular biology* **27**, 8318-8329, doi:10.1128/MCB.01209-07 (2007).

32 Mears, A. J. *et al.* Nrl is required for rod photoreceptor development. *Nature genetics* **29**, 447-452, doi:10.1038/ng774 (2001).

33 Brooks, M. J., Rajasimha, H. K., Roger, J. E. & Swaroop, A. Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes. *Molecular vision* **17**, 3034-3054 (2011).

34 Kim, J. W. *et al.* NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors. *Cell reports* **17**, 2460-2473, doi:10.1016/j.celrep.2016.10.074 (2016).

35 Nadadur, R. D. *et al.* Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Sci Transl Med* **8**, 354ra115, doi:10.1126/scitranslmed.aaf4891 (2016).

36 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

37 Chen, X. *et al.* ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods* **13**, 1013-1020, doi:10.1038/nmeth.4031 (2016).

38 Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**, 207-223, doi:10.1038/nrg.2016.4 (2016).

39 Oh, E. C. *et al.* Transformation of cone precursors to functional rod photoreceptors by bZIP transcription factor NRL. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1679-1684, doi:10.1073/pnas.0605934104 (2007).

40 Yoshida, S. *et al.* Expression profiling of the developing and mature Nrl-/- mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl. *Human molecular genetics* **13**, 1487-1503, doi:10.1093/hmg/ddh160 (2004).

41 Akimoto, M. *et al.* Targeting of GFP to newborn rods by Nrl promoter and temporal expression profiling of flow-sorted photoreceptors. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3890-3895, doi:10.1073/pnas.0508214103 (2006).

42 Livesey, F. J., Furukawa, T., Steffen, M. A., Church, G. M. & Cepko, C. L. Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. *Current biology : CB* **10**, 301-310 (2000).

43 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

44 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

45 Briata, P., Ilengo, C., Bobola, N. & Corte, G. Binding properties of the human homeodomain protein OTX2 to a DNA target sequence. *FEBS letters* **445**, 160-164 (1999).

46 Mouse, E. C. *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* **13**, 418, doi:10.1186/gb-2012-13-8-418 (2012).

47 Emerson, M. M. & Cepko, C. L. Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. *Developmental biology* **360**, 241-255, doi:10.1016/j.ydbio.2011.09.012 (2011).

48 Billings, N. A., Emerson, M. M. & Cepko, C. L. Analysis of thyroid response element activity during retinal development. *PLoS One* **5**, e13739, doi:10.1371/journal.pone.0013739 (2010).

49 Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16-22, doi:10.1073/pnas.2235688100 (2004).

50 Sapkota, D. & Mu, X. Onecut transcription factors in retinal development and maintenance. *Neural Regen Res* **10**, 899-900, doi:10.4103/1673-5374.158350 (2015).

51 Roberts, M. R., Hendrickson, A., McGuire, C. R. & Reh, T. A. Retinoid X receptor (gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of

the developing mouse retina. *Investigative ophthalmology & visual science* **46**, 2897-2904, doi:10.1167/iovs.05-0093 (2005).

52 Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic acids research* **35**, D88-92, doi:10.1093/nar/gkl822 (2007).

53 Zentner, G. E. & Scacheri, P. C. The chromatin fingerprint of gene enhancer elements. *J Biol Chem* **287**, 30888-30896, doi:10.1074/jbc.R111.296491 (2012).

54 Wang, S., Sengel, C., Emerson, M. M. & Cepko, C. L. A gene regulatory network controls the binary fate decision of rod and bipolar cells in the vertebrate retina. *Developmental cell* **30**, 513-527, doi:10.1016/j.devcel.2014.07.018 (2014).

55 Rodgers, H. M., Belcastro, M., Sokolov, M. & Mathers, P. H. Embryonic markers of cone differentiation. *Mol Vis* **22**, 1455-1467 (2016).

56 de Melo, J., Peng, G. H., Chen, S. & Blackshaw, S. The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons. *Development* **138**, 2325-2336, doi:10.1242/dev.061846 (2011).

57 Chang, B. *et al.* Cone photoreceptor function loss-3, a novel mouse model of achromatopsia due to a mutation in Gnat2. *Investigative ophthalmology & visual science* **47**, 5017-5021, doi:10.1167/iovs.05-1468 (2006).

58 Li, X., Perissi, V., Liu, F., Rose, D. W. & Rosenfeld, M. G. Tissue-specific regulation of retinal and pituitary precursor cell proliferation. *Science* **297**, 1180-1183, doi:10.1126/science.1073263 (2002).

59 Montana, C. L. *et al.* Reprogramming of adult rod photoreceptors prevents retinal degeneration. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1732-1737, doi:10.1073/pnas.1214387110 (2013).

60 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

61 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

62 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

63 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258, doi:10.1093/bioinformatics/btl567 (2007).

64 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

65 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

66 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

67 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

888 68 Cherry, T. J. *et al.* NeuroD factors regulate cell fate and neurite stratification in
889 the developing retina. *J Neurosci* **31**, 7365-7379, doi:10.1523/JNEUROSCI.2555-
890 10.2011 (2011).
891 69 Matsuda, T. & Cepko, C. L. Controlled expression of transgenes introduced by in
892 vivo electroporation. *Proceedings of the National Academy of Sciences of the*
893 *United States of America* **104**, 1027-1032, doi:10.1073/pnas.0610155104 (2007).
894 70 Montana, C. L., Myers, C. A. & Corbo, J. C. Quantifying the activity of cis-
895 regulatory elements in the mouse retina by explant electroporation. *J Vis Exp*,
896 doi:10.3791/2821 (2011).
897

Figure 1: Transcriptional profiling identifies *Nrl*-depedent coding and non-coding RNAs

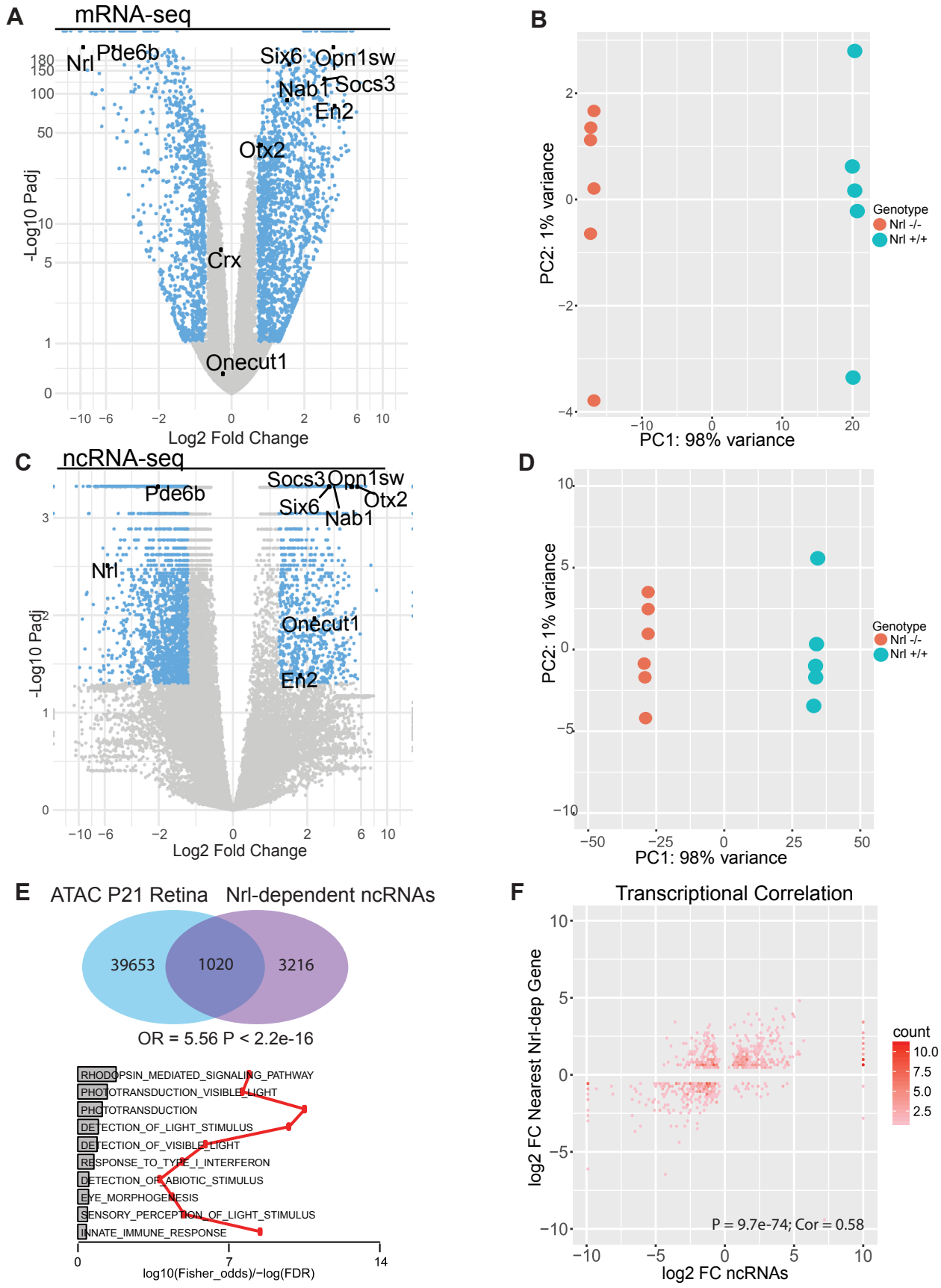


Figure 2: *Nrl*-activated and repressed ncRNAs identify cell-type specific regulatory programs

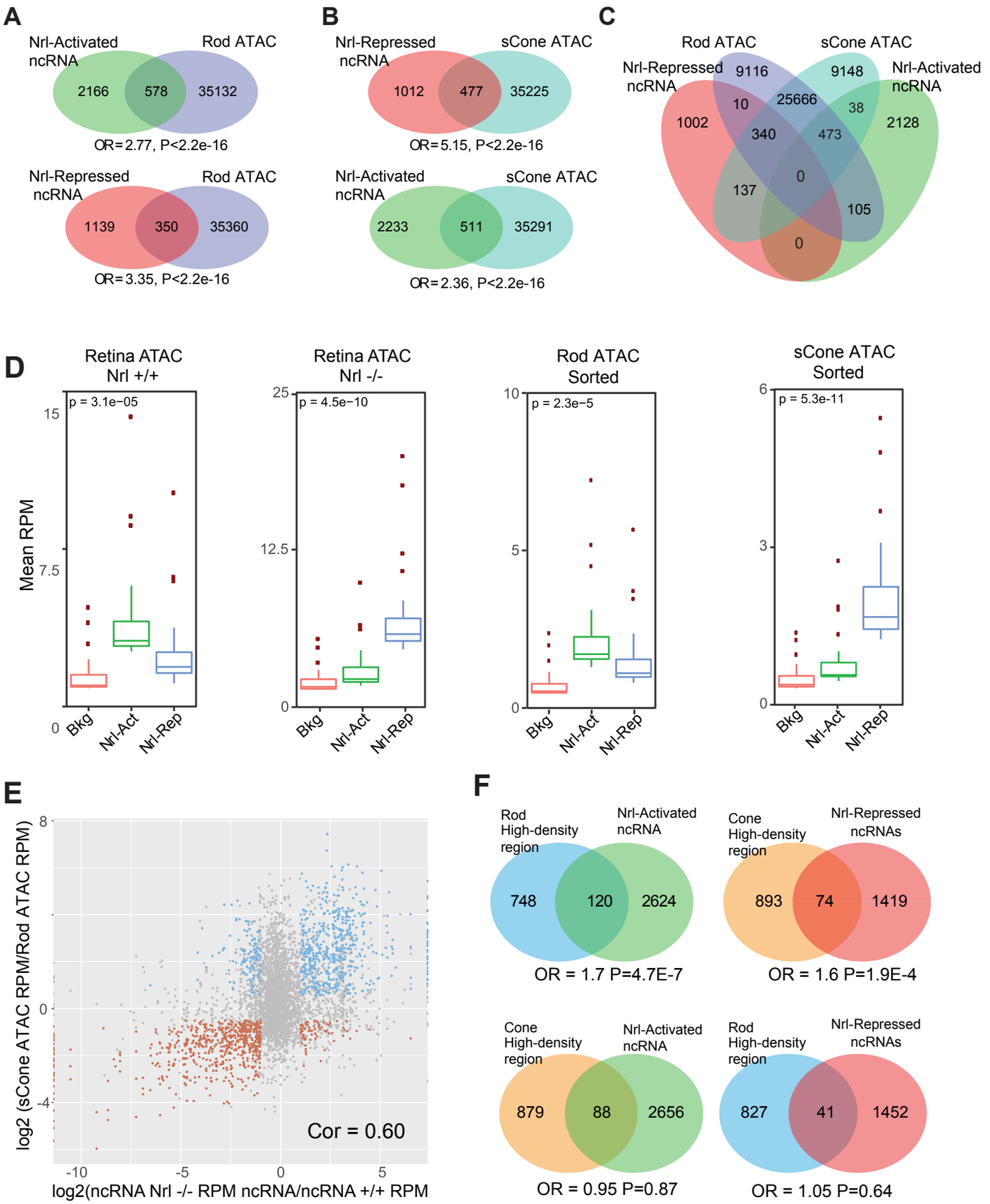
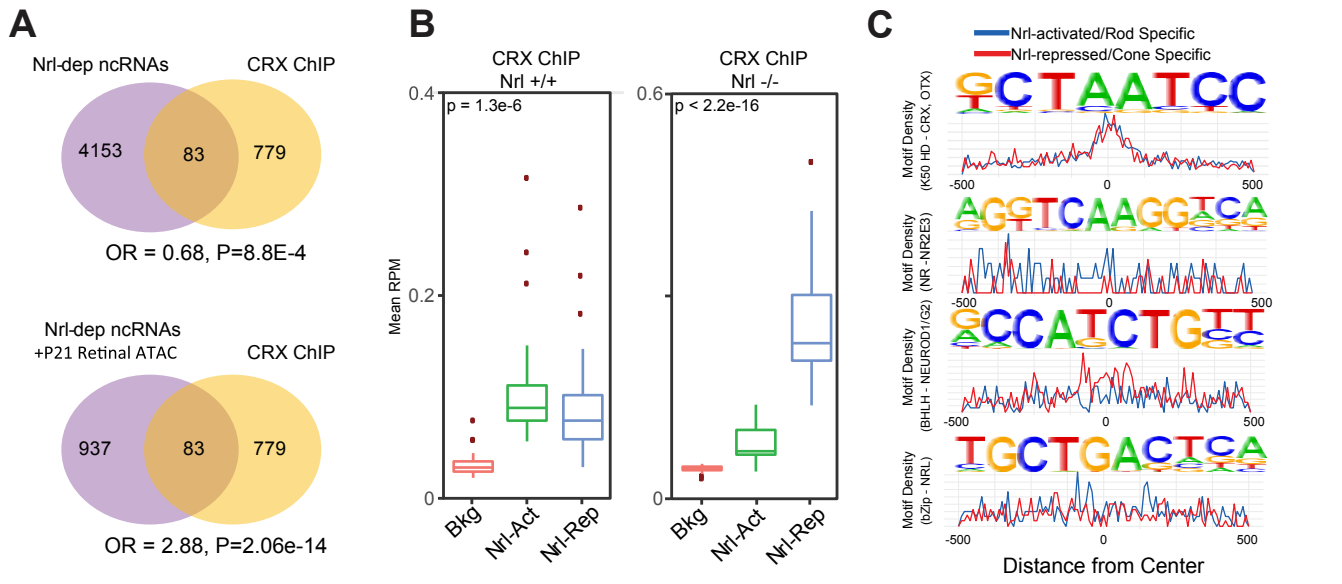


Figure 3. *Nrl*-dependent ncRNAs define cell fate specific regulatory elements



D

	NR (RORA, NR2E3, ESRRB, RXRG)	1e-37
	T-Box (TBX 15)	1e-37
	ZBTB (HIC, THAP)	1e-31
	HMG-box (SOX4)	1e-28
	ZF (Znf416)	1e-27
	MADS (MEF2C, MEF2D)	1e-25
	KLF (KLF1)	1e-25
	NK-2 homeobox (NKX2-3)	1e-25
	PITX homeobox(PITX1)	1e-24
	RHOX homeobox (RHOXF1)	1e-24

E

	ZF (Znf263)	1e-33
	bHLH (NEUROD1, NEUROD2)	1e-31
	bZIP/G-box (AP-1)	1e-31
	Forkhead (FOX, P1)	1e-31
	Fos (FOSB, FOSL2)	1e-27
	EYS (EYS2, ETV5, ETV6)	1e-27
	NF-Y (NFYA, HAP2)	1e-25
	ARID (AT-rich, ARID3A)	1e-22
	Q50 Homeobox (HOXD10, RAX)	1e-19
	NR/ZF (PGR, GATA, GRE)	1e-19

Figure 4: *Nrl*-repressed ncRNAs identify cone-specific regulatory elements with regulatory activity *ex vivo*

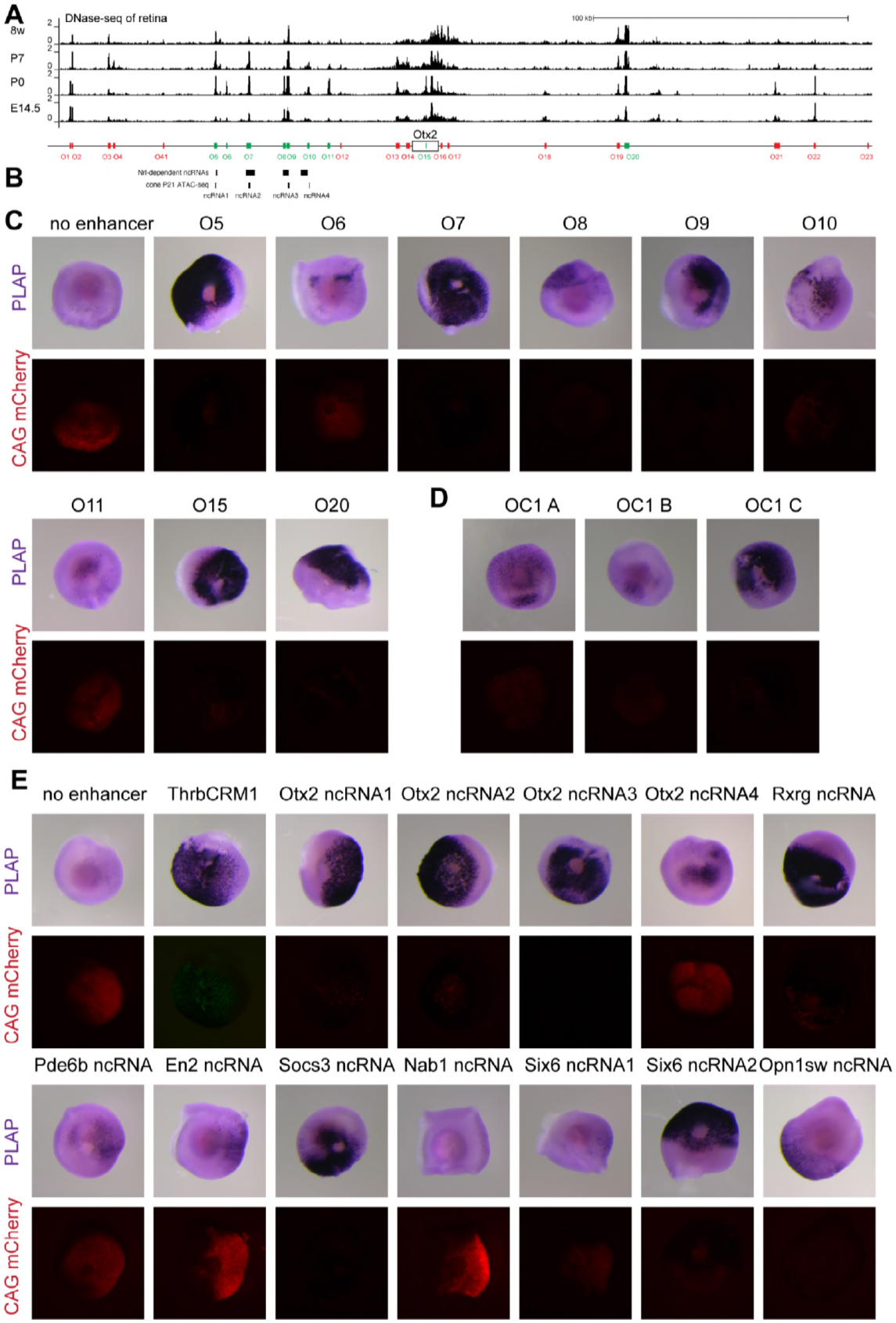


Figure 5: *Nrl*-repressed ncRNAs identify functional cone-specific regulatory elements *in vivo*

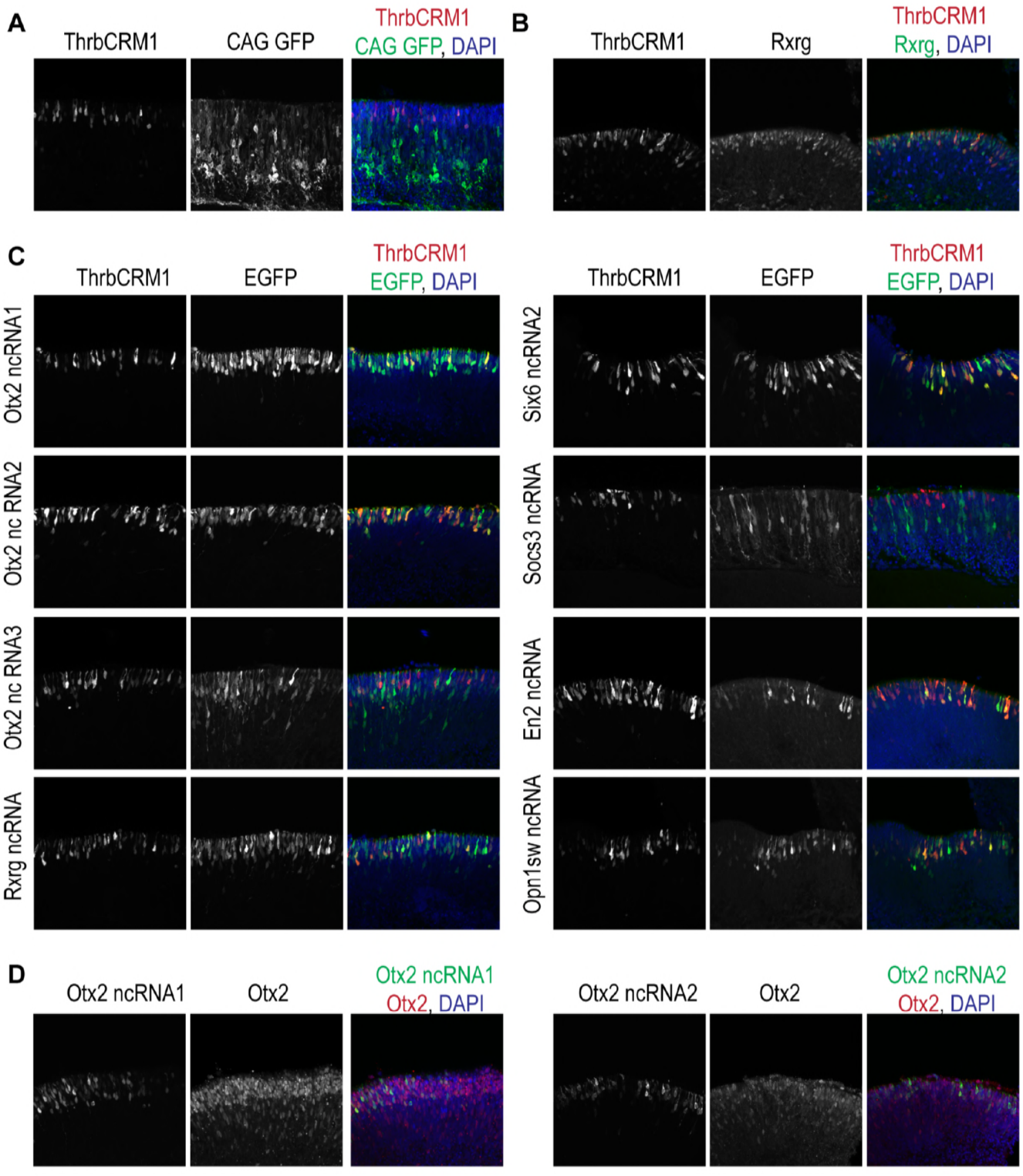


Figure 1S. Processing pipelines for mRNA-seq, ncRNA-seq and ATAC-seq