

RESEARCH ARTICLE

Identification and prediction of alternative transcription start sites that generate rod photoreceptor-specific transcripts from ubiquitously expressed genes

Evgenya Y. Popova^{1,2}, Anna C. Salzberg³, Chen Yang¹, Samuel Shao-Min Zhang^{1,2}, Colin J. Barnstable^{1,2*}

1 Department of Neural and Behavioral Sciences, Penn State University, College of Medicine, Hershey, Pennsylvania, United States of America, **2** Penn State Hershey Eye Center, Hershey, Pennsylvania, United States of America, **3** Bioinformatics Core, Penn State University College of Medicine, Hershey, Pennsylvania, United States of America

* cbarnstable@psu.edu



OPEN ACCESS

Citation: Popova EY, Salzberg AC, Yang C, Zhang SS-M, Barnstable CJ (2017) Identification and prediction of alternative transcription start sites that generate rod photoreceptor-specific transcripts from ubiquitously expressed genes. PLoS ONE 12 (6): e0179230. <https://doi.org/10.1371/journal.pone.0179230>

Editor: Alfred S Lewin, University of Florida, UNITED STATES

Received: March 9, 2017

Accepted: May 25, 2017

Published: June 22, 2017

Copyright: © 2017 Popova et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Support was provided by Macula Vision Foundation for CJB. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Transcriptome complexity is substantially increased by the use of multiple transcription start sites for a given gene. By utilizing a rod photoreceptor-specific chromatin signature, and the RefSeq database of established transcription start sites, we have identified essentially all known rod photoreceptor genes as well as a group of novel genes that have a high probability of being expressed in rod photoreceptors. Approximately half of these novel rod genes are transcribed into multiple mRNA and/or protein isoforms through alternative transcriptional start sites (ATSS), only one of which has a rod-specific epigenetic signature and gives rise to a rod transcript. This suggests that, during retina development, some genes use ATSS to regulate cell type and temporal specificity, effectively generating a rod transcript from otherwise ubiquitously expressed genes. Biological confirmation of the relationship between epigenetic signatures and gene expression, as well as comparison of our genome-wide chromatin signature maps with available data sets for retina, namely a ChIP-on-Chip study of Polymerase-II (Pol-II) binding sites, ChIP-Seq studies for NRL- and CRX- binding sites and DHS (University of Washington data, available on UCSC mouse Genome Browser as a part of ENCODE project) fully support our hypothesis and together accurately identify and predict an array of new rod transcripts. The same approach was used to identify a number of TSS that are not currently in RefSeq. Biological confirmation of the use of some of these TSS suggests that this method will be valuable for exploring the range of transcriptional complexity in many tissues. Comparison of mouse and human genome-wide data indicates that most of these alternate TSS appear to be present in both species, indicating that our approach can be useful for identification of regulatory regions that might play a role in human retinal disease.

Introduction

The most basic identifier of a cell or tissue type is given by its transcriptome. However, the actual diversity of the proteome of a cell is much greater than the 20,000 to 30,000 protein-coding genes found in the mammalian genome. This diversity is achieved by a variety of mechanisms including alternative initiation and termination of transcripts, as well as posttranscriptional processes such as alternative splicing. Although alternative splicing plays a key role in many tissues, particularly the nervous system, recent studies have demonstrated that alternative initiation and termination have a greater contribution to transcriptome diversity [1–3]. Approximately half of mouse and human genes have alternative TSS and promoters [4–6]. Though some of the alternative TSS generate novel proteins, many result in identical open reading frames but different 5'-untranslated regions. Since the 5'-regions of mRNA contain binding sites for regulatory proteins, the use of alternative TSS can provide different pathways for controlling mRNA stability and translation [7–9]. More importantly, the different TSS have different promoters and thus the potential for different sets of transcription factors to regulate the temporal and spatial patterns of transcription. As well as different core promoter structures, alternative TSS can involve different chromatin/epigenetic state and histone modifications. Recent genome-wide studies that mapped different chromatin/epigenetic features found that alternative TSS/promoters have distinctive chromatin states [1, 10, 11]. Conversely, chromatin/epigenetic signatures can be used for prediction of new TSS and their usage in specific cell type or tissue [5, 12–14].

We have been using the mouse retina as a model system to understand the regulation of gene expression during development of different cell types from common progenitors. Our current understanding of the retina transcriptome has been derived initially from studies based on expression arrays [15–23] that lack information about transcription from alternative TSS, and more recently from RNA-seq data that still counts expression per gene, not for separate gene transcripts [24], even though the RNA-seq method could provide expression per isoform as well as per gene. Methods based on cap analysis of gene expression (CAGE) utilize trapping of RNA cap on 5' end of transcript have been used to locate TSS in the genome [25–27]. Accurate positional information on TSS has been collected genome-wide for several tissues and cell types and is available in the Database of TSS (DBTSS) [26, 28, 29]. One CAGE based study for retina, was recently published, but focused on axonal injury and lacked information on the developmental aspects of retina [30].

In an alternative approach, chromatin features and epigenetic signatures can be used as predictors for gene expression status and expression levels [14]. Recently we used ChIP-seq to map the distribution of two important histone H3 modifications, H3K4me2 and H3K27me3, over the whole genome at multiple time points during late mouse retina development [31]. We merged these data with our previous retina developmental expression profiles [16] to define a striking epigenetic signature in genes known to generate a rod photoreceptor-specific transcript (termed “known rod genes” in this manuscript). We then used this signature in an unbiased search of the whole genome and identified a total of 107 genes: 36 known rod genes as well as a group of 71 novel genes (termed “new rod genes” in this manuscript) that we hypothesized have a high probability of generating transcripts in rod photoreceptors that are not found in other cell types.

In the present manuscript we have shown that many of these novel genes are transcribed into multiple mRNA or protein isoforms, using one TSS to generate a transcript in rod photoreceptors and another to generate a transcript in a broader array of cell types. We have also compared our genome-wide chromatin signature maps with ChIP-Seq studies for binding sites for CRX and NRL, two of the best characterized transcription factors for rod photoreceptor genes

[32, 33], as well as available data sets for retina ChIP-on-Chip studies of Polymerase-II (Pol-II) binding sites [34] and DNase I hypersensitive sites (DHS), specific sites in genome where chromatin is more open and accessible (University of Washington data, available on UCSC mouse Genome Browser as a part of ENCODE project)[35]. This comparison fully supports our hypothesis that cell type restricted TSS can be determined and predicted by specific epigenetic signatures.

Materials and methods

Reagents

All chemicals were purchased from Fisher Scientific (Pittsburgh, PA), unless otherwise noted.

Animals

All animal experiments were conducted in accordance with NIH guidelines and were approved by the Animal Care and Use Committee of Pennsylvania State University School of Medicine (Protocol # 46432). C57BL/6j and C3H/HeJ*Pde6b^{rd1}* mice were purchased from the Jackson Laboratory (Bar Harbor, ME) and bred in the animal facility of Penn State College of Medicine. Mice were housed in standard cages with a 12h light/12h dark cycle. Rodent chow and water were provided ad libitum.

Genome-wide databases

Data sets were collected from the NCBI GEO data repository and visualized in the UCSC Genome Browser (www.genome.ucsc.edu). Data sets for H3K4me2 CHIP-Seq were from mouse retina at E17, PN1 PN7, PN15 of wild type (WT) mice and from retina of mutant RD1 PN30 (GSE38500, [31]); mouse early and late erythroblast cells from embryonic liver (GSE27893, [36]); mouse NP (neuronal progenitor) and WB (whole brain) (GSE11172, [37]).

For tag distribution around TSS we used RefSeq database of established transcription start sites [38] and the number of reads were calculated in the window TSS \pm 1000bp.

The CRX-binding site ChIP-Seq data set was from WT retina at PN56 (GSE20012, [33]). The number of reads was calculated in a window of \pm 1000bp around the TSS. We classified TSS according to CRX binding at the cluster of genes of interest. Binding was clearly bimodal with minimum binding more than 300 reads, 0 reads was classified as no binding and everything above this as positive binding (Fig A in [S1 File](#)).

Similarly, we studied NRL binding to TSS using a published Chip-Seq data set from WT retina PN28 [32]. The number of reads was calculated in a window of \pm 1000bp around the TSS. Binding was clearly bimodal and TSS with 0 reads was classified as no binding and everything above this as positive binding (Fig B in [S1 File](#)).

The ChIP-on-Chip dataset for Polymerase Pol-II binding sites was obtained from WT retina at PN2 and PN25 (GSE19999, [34]). We estimated stage-specific PolII occupancy in \pm 1000bp window of each TSS for genes in [S1 Table](#). We then calculated level of PolII-binding changes during development by subtracting PolII occupancy at PN2 from PolII occupancy at PN25 (Δ PolII). We divided TSS accordingly to developmental changes in PolII binding in three groups: PolII is not binding to TSS (Δ PolII = 0), PolII binding is low but up-regulated during development ($0 < \Delta$ PolII < 182, where 182 = two S.D. + mean), PolII binding is highly up-regulated during development (Δ PolII > 182).

Mouse DNaseI Hyper Sensitive sites were collected from the ENCODE project: mouse retina PN1(1D), PN7 (1W), PN56 (8W) (GSM1014188; GSM1014198; GSM1014175)[35]; mouse whole brain from E14.5, E18.5 and PN56 (GSM1014197, GSM1014184, GSM1014151). Each

dataset above was intersected with each isoform's TSS +/- 1000 bp, with the score contribution of a dataset entry being set to the number of intersecting bps times the score of the entry. We calculated ratio of DHS changes during development by dividing of DHS occupancy at PN56 and PN1 (PN56/PN1). We then divided TSS according to developmental changes in DHS occupancy in four groups: DHS is down regulated during development (DHS PN56/PN1 < 1), DHS is up regulated during development ($1 < \text{DHS PN56/PN1} < 1.35$), DHS is strongly up regulated during development (DHS PN56/PN1 > 1.35) (where 1.35 = two S.D+ mean).

Developmental profiles of gene expression were from reanalyzed data of microarrays previously described [16]. Expression was normalized to a pool of equal amounts of all seven ages tested, resulting in a maximum sevenfold change in expression.

Data access

We have made data viewable on UCSC mouse Genome Browser mm9 with our data sets for H3K4me2/ H3K27me3 profiles, as well publicly available data sets for DHS (ENCODE project, [35]), PolII- and CRX binding sites [33, 34] (Barnstable lab retina epigenome).

Cluster analysis

Hierarchical cluster analysis was performed with Gene Cluster 3.0[39] and Java Tree View was used for clusters visualization or with the heatmap.2 function of the gplots R package, using $(1 - \text{correlation})/2$ as the distance and average as the clustering method. The Z scaling was obtained from the genefilter v1.46.1 R package.

For cluster comparisons we use IBM developed analysis for clusters (https://www.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/alg_cluster_evaluation_goodness.htm). In particular, the 'silhouette coefficient' combines the concepts of cluster cohesion (favoring models which contain tightly cohesive clusters) and cluster separation (favoring models which contain highly separated clusters). It can be used to evaluate individual objects, clusters, and models. Average Silhouette coefficient: $(B-A)/\max(A,B)$ where A is the average distance from the case to every other case assigned to the same cluster and B is the minimal average distance from the case to cases of a different cluster across all clusters. An average silhouette coefficient greater than 0.5 indicates reasonable partitioning of data while average silhouette coefficient less than 0.2 means that the data do not exhibit cluster structure.

Tissue collection

Animals were euthanized in Euthanex Auto CO₂ System (www.euthanex.com). Whole retinas were isolated from animals by quickly removing the sclera and most of the retinal pigmented epithelium (RPE) layer in PBS. Spleen, brain, lung and liver tissue were quickly isolated from mouse and cleaned of connective tissue. Immediately after isolation, tissue was flash frozen in liquid nitrogen and stored in -80°C. 30 mg of tissue or 2 retinas were used for each RNA extraction.

RNA extraction and cDNA preparation

RNA extraction and purification followed the manufacturer's protocol from RNeasy Mini Kit and RNA shredder (Qiagen). Buffer RLT was added to the frozen tissue and spleen, brain, lung or liver tissue was disrupted in a teflon-glass homogenizer and retinas were homogenized by pipet trituration at room temperature according to the Kit protocol. Final RNA concentrations were determined spectrophotometrically using a GeneSpect III (Hitachi Tokyo, Japan). cDNA

was synthesized with SuperScriptII First-Strand Synthesis System kit according to manufacturer's protocol (Invitrogen, Carlsbad, California).

RT-PCR

Primers were designed and purchased from Integrated DNA Technologies (IDT). The sequence information listed in [S1 Table](#). For quantitative real-time PCR we used 2x iQ-SYBR Green PCR supermix from Bio-Rad. Samples in triplicate were run on iQ5 Multicolor Real Time PCR Detection System (Bio-Rad).

Statistical analysis

Statistical analyses were performed using the Excel or GraphPad Prism software. Student's t-test (two-tailed, unpaired) was used to compare two groups.

Results

Epigenetic signatures indicate tissue-specific use of TSSs of genes expressed in many tissues

Among the 107 genes with a distinct rod photoreceptor epigenetic signature identified in our previous study were a group of 30 with a single TSS that have previously been shown to be expressed exclusively in rod photoreceptors [31]. Throughout this earlier study, this set of 30 known rod genes was used as a training set to help define the characteristics of other transcripts selectively expressed in rod photoreceptors. A wide range of databases and publications have documented that 102 of the 107 genes identified in our earlier study show expression in the eye; there is no published information on the other 5. In addition to the published data, a group chosen at random showed an increase in retinal expression that matched the time course of rod differentiation, and decreased expression in the RD1 mutant at a time when all the rods have degenerated [31], thus supporting the idea that these genes are all expressed in rod photoreceptors.

Of the 107 genes with a rod specific epigenetic signature, 72 (67%) have a single identified TSS listed in the RefSeq database. Of the other 35 genes, 27 have two identified TSS and 8 have 3 TSS, for a total of 78 TSS ([Fig 1A](#), [S2 Table](#)). These alternative TSSs lead 25 of the genes to produce different isoforms of the protein and 10 of the genes to produce different mRNAs but the same protein isoform ([S2 Table](#)).

To explore further the relationship of the epigenetic signatures around TSS (+/-1000bp) with other characteristic markers of gene activity we extracted data for several chromatin and transcriptional features for the 107 genes defined by all their TSS. In addition to H3K4me2 occupancy (at ages E17, PN1, PN7, PN15 and RD1), we used data on binding of the two best characterized rod photoreceptor transcription factors CRX (at age PN56) and NRL (at age PN28). We also used the developmental appearance of DHS (at ages PN1(1D), PN7(1W), PN56(8W)) and the general transcription marker of PolII-binding (at ages PN2 and PN25) ([S3 Table](#)). For each gene discussed below we provide all the marker data but discuss the general applicability of each marker in separate sections below. RD1 mouse PN30 retina, an age where the retina has lost all rod photoreceptors, was used as negative control. Bipolar cells undergo extensive remodeling during neurodegeneration and in RD1 mice (review in[40]), but expression of rod-specific genes as *Rho*, *Sag* and *Rorb* show substantial down regulation in gene expression arrays that compare mouse retina at 3 developmental points from RD1 mutant and wild type retina [41].

We first performed a hierarchical clustering analysis using genome-wide data for all of these markers available for mouse retina ([Fig 1B](#)). All TSS fell into 1 of 2 very distinct clusters.

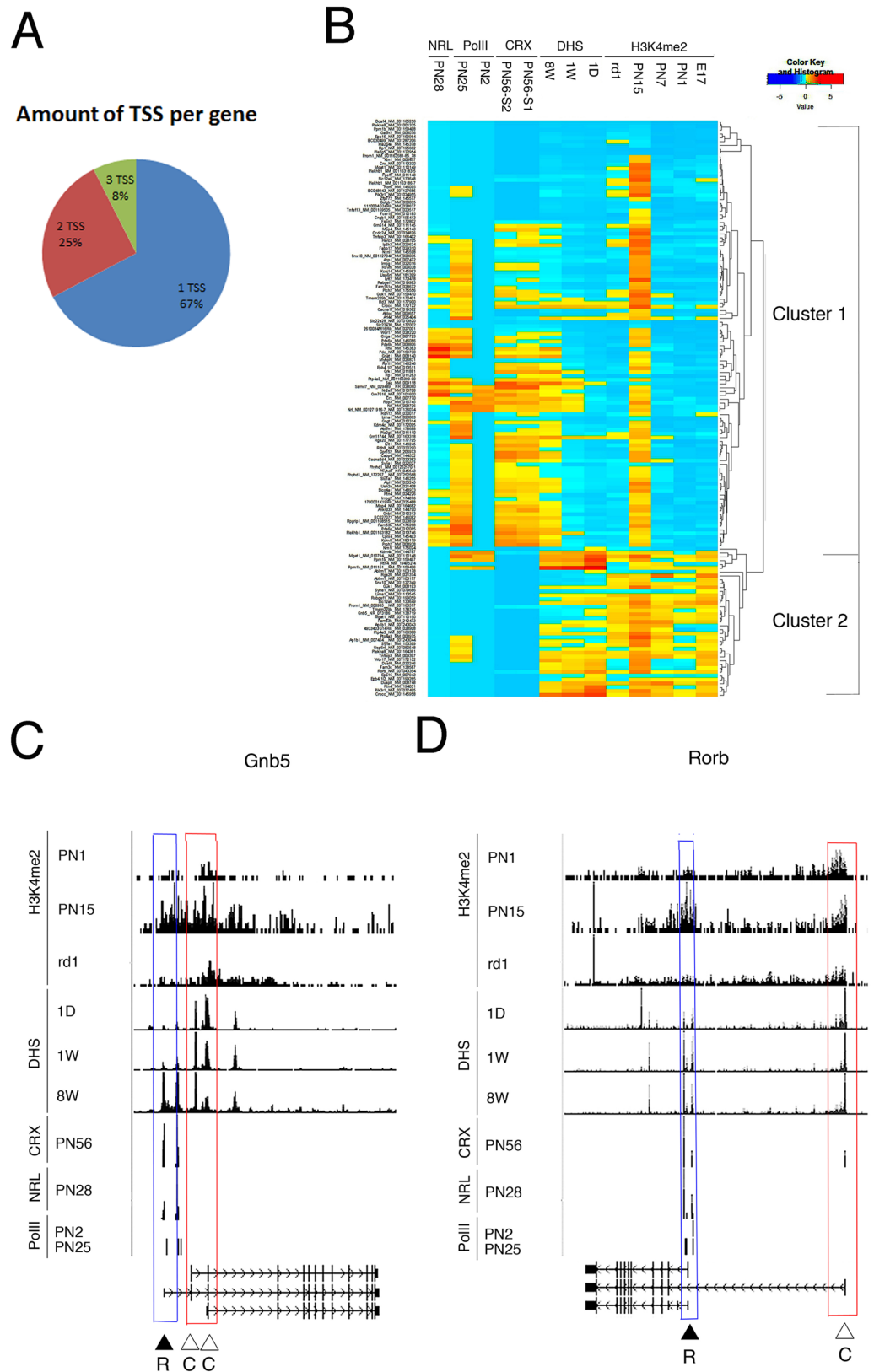


Fig 1. Epigenetic signature predicts employment of ATSS of ubiquitous gene in tissue-specific manner.
A. Pie chart presentation for number of TSS for different groups of rod-specific genes. **B.** All TSS for rod-specific

genes were clustered based on chromatin signature around TSS \pm 1000bp for following features: developmental changes in H3K4me2 occupancy (E17, PN1, PN7, PN15 and RD1 PN30), CRX-binding (PN56), NRL-binding (PN28), developmental changes in DHS (1D, 1W, 8W), developmental changes in PolII (PN2, PN25) as clustering criteria (see [methods](#) for details). **C, D.** Combined genome-wide tracks of chromatin features, as in **B** for *Gnb5* and *Rorb* genes. Common TSS is depicted as C/ red box, rod TSS—as R/ blue box.

<https://doi.org/10.1371/journal.pone.0179230.g001>

Cluster 1 contains the TSS from all 72 genes with a single TSS, including all of our training set of known rod genes, and one (or more) TSS of the genes with multiple TSS ([S4 Table](#)). Cluster 1 TSS can be defined by the combination of H3K4me2 accumulation during development (less H3K4me2 in retina of RD1 mice), PolII-binding and DHS increase later in development and binding of CRX and NRL. We refer to these as *rod TSS* throughout this paper.

Cluster 2 defines the additional TSS from the 35 genes with multiple TSS ([S4 Table](#)). These TSS show a constant accumulation of H3K4me2 and constant DHS during development, as well as no NRL or CRX-binding, suggesting that they could be used constitutively in retina and possible, other cell types. We refer to these as *common TSS* throughout this paper.

Rods are already postmitotic at PN7, but transcription of the large group of later-expressed true rod-specific gene (like Rho, Pde6b) starts at PN7 and then dramatically increases in later stages from PN15 to adulthood (see Fig 1A in [\[31\]](#) and Fig 2D in [\[16\]](#)). Because of this, neither H3K4me2, nor DHS at PN7 show big accumulation at TSS for well-established rod-specific genes. The same is true for our newly predicted rod TSS. As epigenetic profiles for later-expressed genes are very similar at PN1 and E17, and PN7 represented a stage of transition, to simplify figures and further analysis we used PN1 comparing it with PN15 for H3K4me2, and PN1 (1D) and PN56 (8W) for DHS.

To test whether the different TSS environments reflect cell type and temporal specificity we examined two well-characterized rod genes, *Gnb5* and *Rorb*, both exist as two protein isoforms where the tissue expression of each is known ([Fig 1C and 1D](#)). *Gnb5* is a member of the G-protein beta subunit family. GNB5 expression is found almost exclusively in the nervous system. The longer form is expressed only in the retina [\[42\]](#), is important for regulating G protein-mediated signaling cascades in photoreception [\[43\]](#), and its TSS (depicted as R in [Fig 1D](#)) has the features of the rod epigenetic signature. The shorter protein form of GNB5 is expressed in brain and brain-derived cell lines [\[44, 45\]](#) and its two TSS (depicted as C in [Fig 1D](#)) show unaltered high H3K4me2 accumulation for all developmental stages. These two common TSS are marked by DHS regions present at birth but the rod TSS develops a DHS in parallel with the time course of rod development. The rod TSS also shows CRX and NRL, binding features absent in the common TSS.

The *Rorb* gene has 2 protein isoforms, transcribed from two TSS where the first exons differ between the two forms. The longer form, ROR β 1 is expressed broadly in the different parts of the brain [\[46\]](#) and amacrine and horizontal cells early in retina development [\[47\]](#). Its common TSS (depicted as C in [Fig 1D](#)), has high levels of H3K4me2 throughout development, no change in H3K4me2 accumulation in the RD1 mutant, a constant DHS even at PN1, and no NRL binding. The shorter form, ROR β 2 is expressed later during retina development and is important for rod differentiation [\[48\]](#). Its rod TSS (depicted as R in [Fig 1D](#)), on the other hand has the rod epigenetic signature, a DHS that develops postnatally and both CRX and NRL binding in TSS \pm 1000bp window that is coincident with the DHS. For both *Gnb5* and *Rorb* the known cell type expression of the protein isoforms can successfully be predicted by the TSS structure.

We next explored the epigenetic signatures around the TSSs of the set of the 35 genes with multiple TSS. During retina development, the TSSs that we predict define a rod transcript accumulate the active epigenetic mark H3K4me2 with the mean PN15/PN1 ratio of 10.1 \pm 2.1

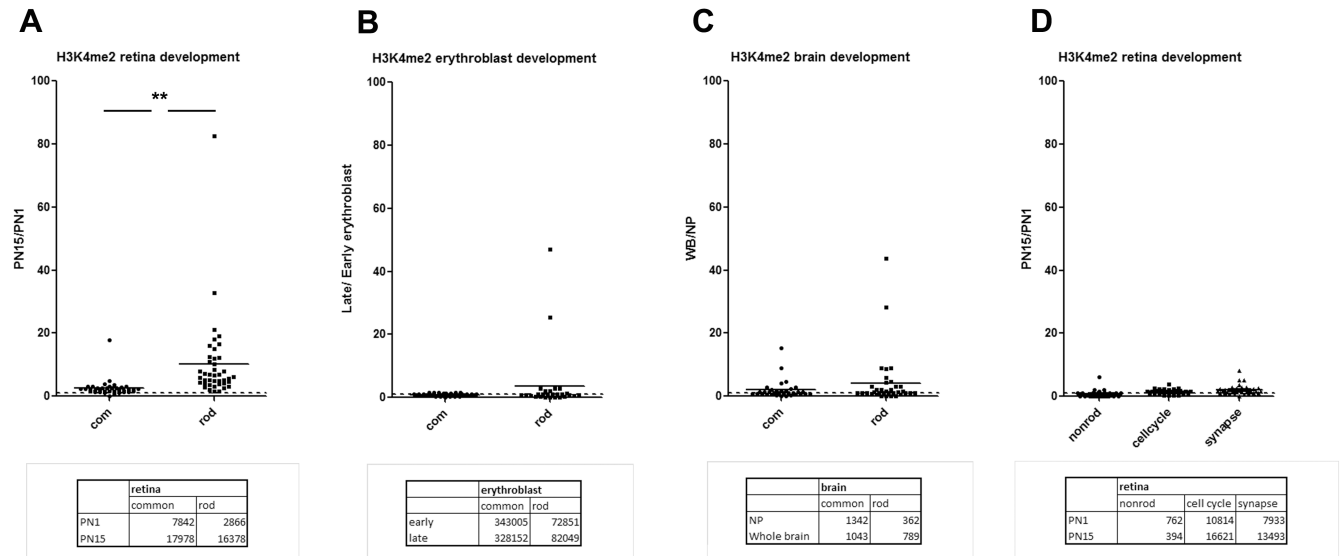


Fig 2. H3K4me2 epigenetic signature predicts tissue and cell specificity of TSS. **A.** Changes of H3K4me2 occupancy around TSS +/-1000bp during mouse retina development as ratio of normalized number of reads at PN15/PN1 for rod and common TSS of retina genes with multiple TSS. At the bottom: average number of H3K4me2 reads accumulation for each group of TSS at PN1 and PN15. **p = 0.0011. **B.** Changes of H3K4me2 occupancy around TSS +/-1000bp during mouse erythropoiesis as ratio of normalized number of reads between late and early erythroblast for rod and common TSS of retina genes with multiple TSS. At the bottom: average number of H3K4me2 reads accumulation for both groups of TSS at early and late stages. **C.** Changes of H3K4me2 occupancy around TSS +/-1000bp during mouse brain development as ratio of normalized number of reads for whole brain and neuronal progenitor for rod and common TSS of retina genes with multiple TSS. At the bottom: average number of H3K4me2 reads accumulation for both groups of TSS at early and late stages. **D.** Changes of H3K4me2 occupancy around TSS +/-1000bp during mouse retina development for TSS of control groups of genes (non-rod, cell-cycle and synapse).

<https://doi.org/10.1371/journal.pone.0179230.g002>

in the region +/- 1000bp around the TSS (Fig 2A). The other TSSs, that we predict produces a tissue common transcript, did not show such a large change in accumulation and had a mean ratio of 2.6 +/- 0.5. The 4 fold higher ratio in the rod TSSs was highly significant (p = 0.0011). We then examined H3K4me2 accumulation at the TSS of the same genes in blood between late and early erythroblast [36] and for brain between whole brain (WB) and neuronal progenitor (NP) [37]. There were no significant developmental differences in H3K4me2 accumulation for either rod or common TSS of the 35 genes during blood (Fig 2B) or brain (Fig 2C) development.

To exclude the possibility that the developmental increase in H3K4me2 at rod TSSs might be a general developmental effect rather than a rod specific effect, we picked approximately 30 genes for three different functional groups: cell cycle and proliferation genes, genes involved in neural processes and synapses, and genes not expressed in retina (S5 Table). None of these genes showed any significant developmental difference in accumulation on H3K4me2 around TSS in retina (Fig 2D).

These data support the idea that for genes that may be expressed in multiple tissues, alternative TSS can be used to generate tissue-specific transcripts and can be recognized by their epigenetic signature.

Biological confirmation of relationship between epigenetic signatures and gene expression

We compared levels of expression from the rod and common TSS of several of the genes with alternative TSS (S2 Table) using qRT-PCR with primers designed to identify the different transcripts from each TSS. Two examples are shown in Fig 3.

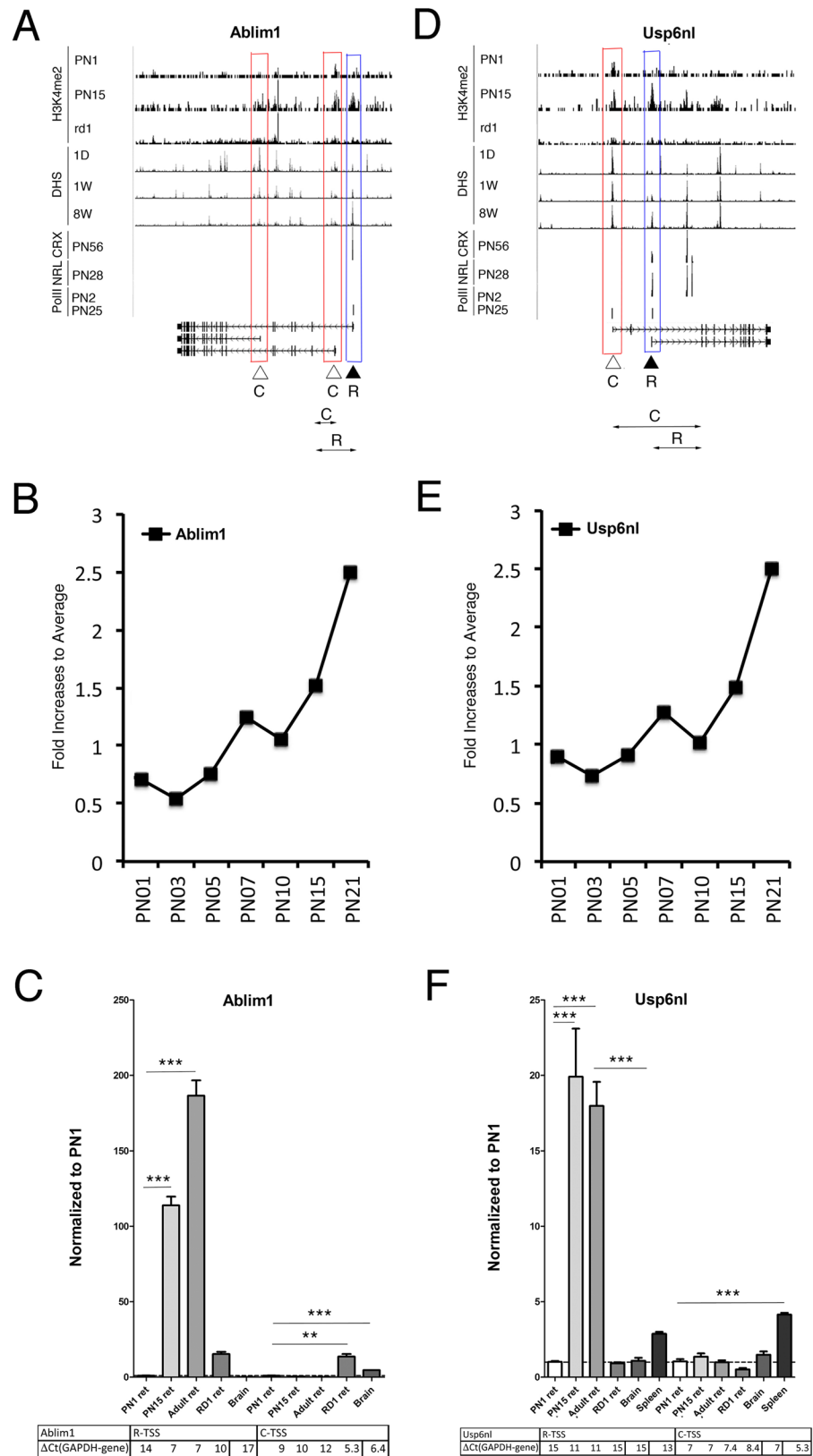


Fig 3. Confirmation of relationship between epigenetic signatures and gene expression. A, D. Combined genome-wide tracks of chromatin features, as in Fig 1B for *Ablim1* (A) and *Usp6nl* (D) genes.

Common TSS is depicted as C/ red box, rod TSS—as R/ blue box. Position of the specific primer sets and PCR product that were used to assess and distinguish TSS-specific gene expression by RT-PCR depicted as double arrow below gene maps. **B, E.** Patterns of *Ablim1* and *Usp6nl* genes expression during mouse retina development; reanalyzing of microarray data from [16]. **C, F.** Relative gene expressions from rod and common TSS by RT-PCR with primer pairs depicted at **A and D** for *Ablim1* (**C**) and *Usp6nl* (**F**) for mouse retina samples at PN1, PN15, adult and RD1 mutant, compare with mouse brain and spleen. For comparison, normalized to *Gapdh* delta Ct values for each sample are in table below. Experiments done in duplicates with three technical replicates; ***— $p < 0.0001$.

<https://doi.org/10.1371/journal.pone.0179230.g003>

The *Ablim1* gene has three alternative TSS with all forms having different first exons and different protein sequences (Fig 3A). *Ablim1* transcripts have a widespread tissue distribution and microarray data demonstrate an overall increase in expression during retina maturation (Fig 3B); however, the longest isoform is found exclusively in retina [16, 49]. This longest isoform has all the characteristics of being transcribed from a rod TSS. The other two TSS show a similar H3K4me2 distribution between PN1 and PN15, little change in DHS and no CRX or NRL binding (Fig 3A). We confirmed by RT-PCR that the proposed rod isoform is expressed primarily in postnatal retina, while one of transcripts from common TSS has higher expression in the RD1 mutant that has lost rods (Fig 3C).

The *Usp6nl* gene encodes a GTPase activator that is required for the structural integrity of the Golgi complex. It has two alternative TSS and each form has a different first exon and encodes a different protein (Fig 3D). Previously published expression array data [16] demonstrate an overall increase in *Usp6nl* expression during retina maturation (Fig 3E). Inspection of the TSS regions shows that one of the TSS has the properties of a rod TSS and the other of a common TSS (Fig 3D). RT-PCR amplification with specific primer sets confirmed that the transcript from the rod TSS generates an isoform expressed mostly later in postnatal retina and lost in the RD1 retina, while the other TSS generates a transcript, that shows no difference in expression during retina development and in RD1 (Fig 3F).

Only two of our set of new rod genes, *Guk1* and *Rtn4*, gave anomalous results (Fig 4). Although the epigenetic signature, DHS accumulation, both CRX and NRL binding profiles predict rod specificity for one of the TSS and mRNA for the apparent rod specific form increased during postnatal development, expression was much higher in adult RD1 retinas. Neither TSS properties nor expression of *Guk1* and *Rtn4* genes from the other TSS (Fig 4) were altered during development, as expected from a tissue common TSS.

Association of rod photoreceptor transcription factors CRX and NRL with rod TSS

To examine the reliability of using transcription factors to define cell type specific TSS, we combined our genome-wide chromatin signature maps with the whole-genome ChIP-Seq data for CRX and NRL [32, 33]. We calculated CRX and NRL accumulation in +/- 1000bp window of each TSS for all 107 genes.

Since the amount of CRX binding showed a clear bimodal distribution (Fig A in S1 File), we classified genes as having or not having CRX binding in the vicinity of their TSS. Of the 72 genes with a single TSS 55 (76%) had clear CRX binding (Fig 5A). Of the 35 genes with multiple TSS, 41 TSS matched the rod profile (cluster 1 in Fig 1A) and of these 30 (73%) had clear CRX binding (Fig 5A) similar to rod genes with single TSS. In contrast, the 37 TSS defined as common from the epigenetic profiles, only 8% showed CRX binding. The 11 (27%) rod TSS that appeared to have no CRX binding were examined in more detail. Five of these (*Mgat1*, *Ap1b1*, *Pik3r1*, *Plekha8*, and *Prom1*) did have CRX binding but more than 1000 bp from the TSS.

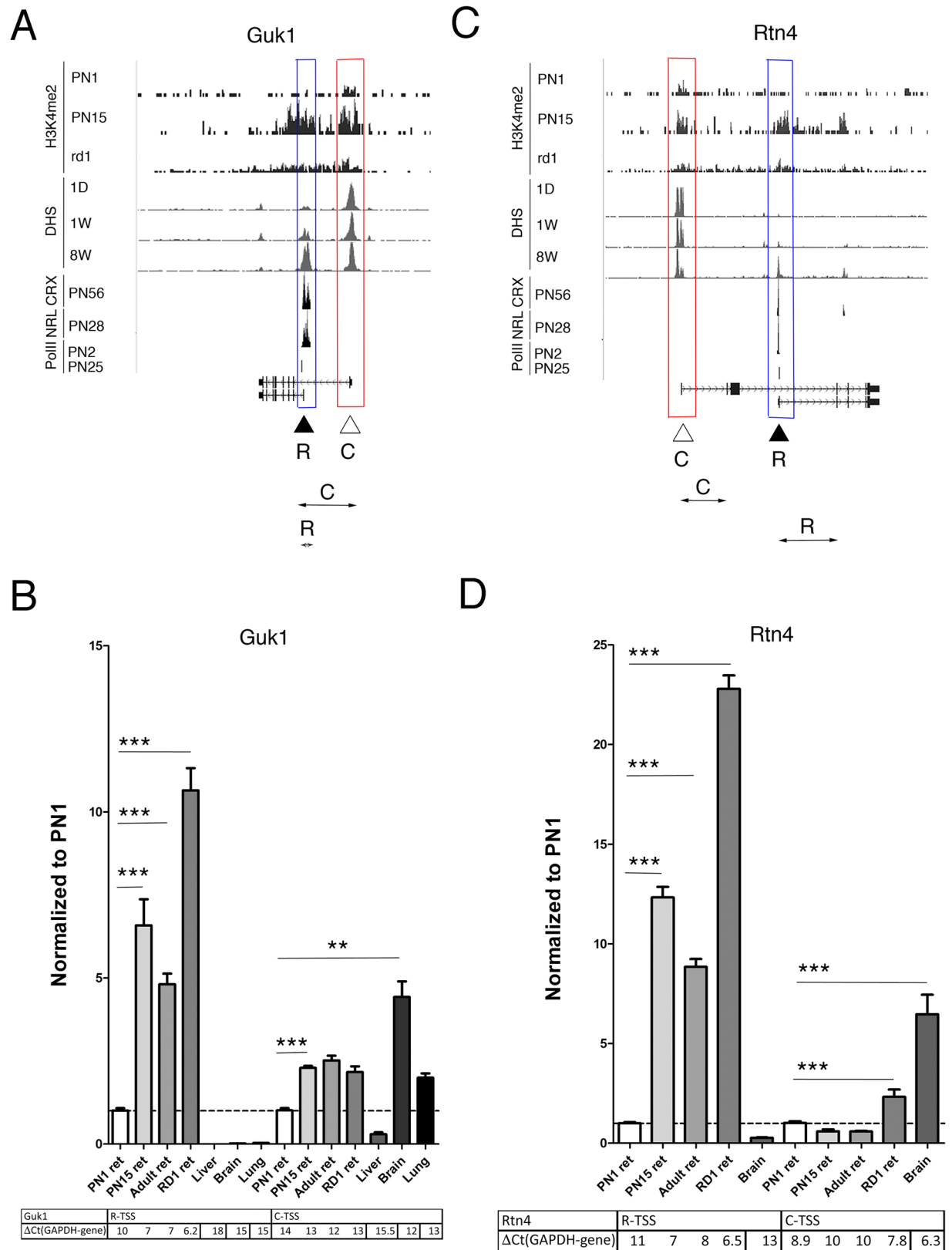


Fig 4. Examples of the genes that use alternative transcription start sites in retina. **A, C.** Combined genome-wide tracks of chromatin features, as in Fig 1B for *Guk1* (A) and *Rtn4* (C) genes. Common TSS is depicted as C/ red box, rod TSS—as R/ blue box.

Position of the specific primer sets and PCR product that were used to assess and distinguish TSS-specific gene expression by RT-PCR depicted as double arrow below gene maps. **B, D.** Relative gene expressions from rod and common TSS by RT-PCR with primer pairs depicted at **A and C** for *Guk1* (**B**) and *Rtn4* (**F**) for mouse retina samples at PN1, PN15, adult and RD1 mutant, compare with mouse brain, liver and lung. For comparison, normalized to *Gapdh* delta Ct values for each sample are in table below. Experiments done in duplicates with three technical replicas; ***— $p < 0.001$; **— $p < 0.0001$.

<https://doi.org/10.1371/journal.pone.0179230.g004>

When the levels of CRX binding were plotted, the mean level for the rod TSS was 22 times that for the common TSS (**Fig 5B**). Analysis of the groups of control non-rod, cell cycle and synapse genes as used earlier showed that their TSS demonstrated almost no CRX binding.

We next analyzed NRL binding to TSS in the same way. This too was bimodal (**Fig B in S1 File**) and we were able to define genes as having or not having NRL binding. Of the 72 genes with a single TSS 44 (61%) had NRL binding (**Fig 5C**). Of the 35 genes with multiple TSS, 23

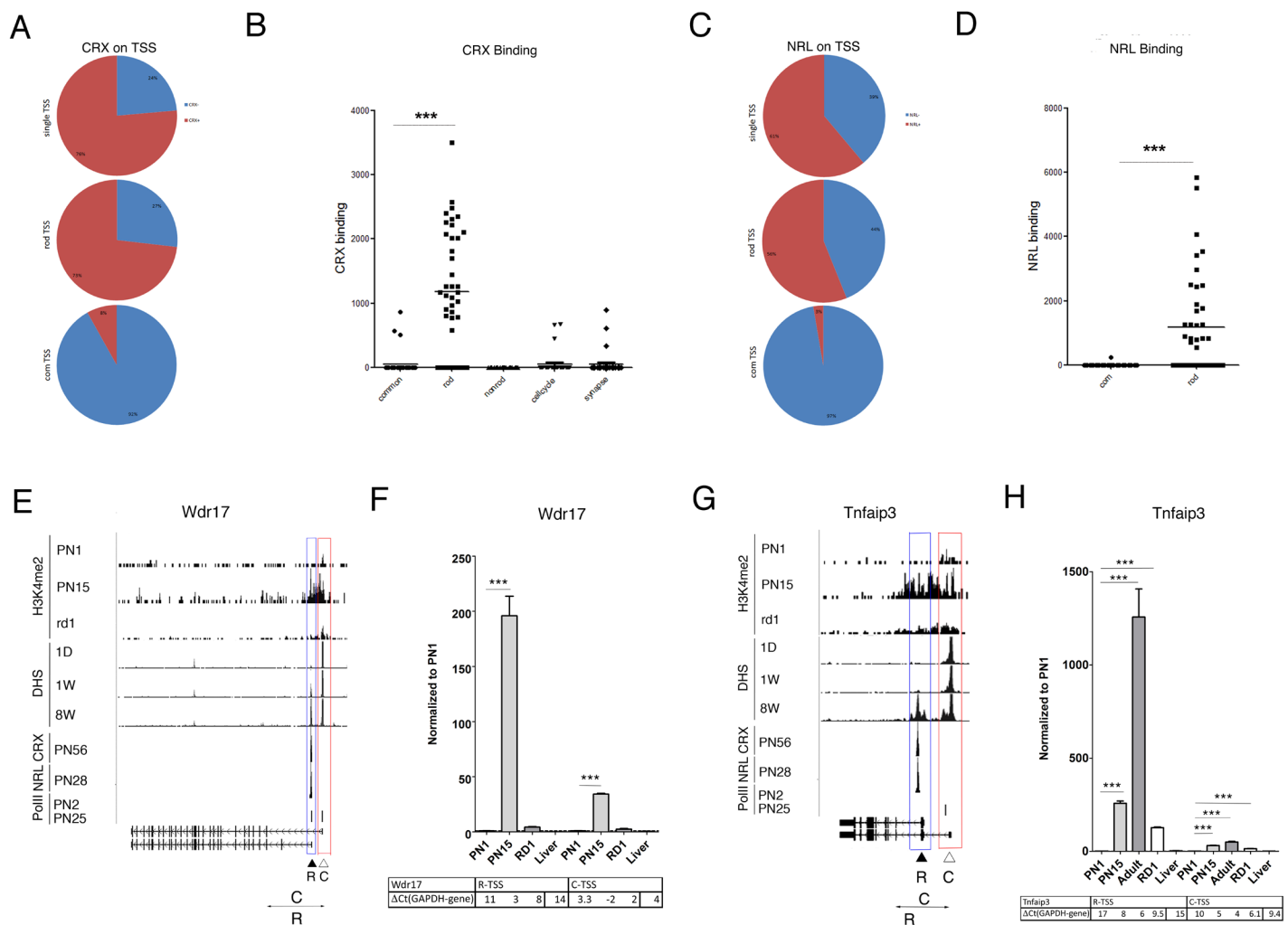


Fig 5. Association of rod photoreceptor transcription factors CRX and NRL with rod TSS. **A, C.** Number of TSS with different CRX (**A**) and NRL (**C**) binding for rod-specific genes with single TSS (upper panel) and for rod (middle panel) or common (bottom panel) TSS of retina genes with multiple TSS. Blue—no TF-binding; red—TF-binding. **B, D.** CRX (**B**) and NRL (**D**) accumulation at TSS (normalized number of reads at TSS \pm 1000bp) for rod and common TSS of retina genes with multiple TSS, compared with control groups of genes (non-rod, cell-cycle and synapse). ***— $p < 0.0001$. **E, G.** Combined genome-wide tracks of chromatin features, as in **Fig 1B** for *Wdr17* (**E**) and *Tnfaip3* (**G**) genes. Common TSS is depicted as C/ red box, rod TSS—as R/ blue box. Position of the specific primer sets and PCR product that were used to assess and distinguish TSS-specific gene expression by RT-PCR depicted as double arrow below gene maps. **F, H.** Relative gene expressions from rod and common TSS by RT-PCR with primer pairs depicted at **E and G** for *Wdr17* (**F**) and *Tnfaip3* (**H**) for mouse retina samples at PN1, PN15, adult and RD1 mutant, compare with mouse liver. For comparison, normalized to *Gapdh* delta Ct values for each sample are in table below. Experiments done in duplicates with three technical replicas; ***— $p < 0.0001$.

<https://doi.org/10.1371/journal.pone.0179230.g005>

(56%) of the 41 rod TSS showed substantial NRL binding, similar to rod genes with single TSS, but only 1 (3%) of the 37 common TSS showed NRL binding (Fig 5C). When the levels of NRL binding were plotted for each TSS, the average for the rod TSS was 172 times greater than for the common TSS (Fig 5D). The TSS of the three control groups of genes had no NRL binding above threshold (data not shown).

Genome-wide tracks for two genes, *Wdr17* (Fig 5E) and *Tnfaip3* (Fig 5G), exemplify these patterns of retinal TF binding at rod TSS. *Wdr17* is a candidate for the RP29 form of Retinitis Pigmentosa [50, 51] and has binding sites for CRX and NRL close to the rod TSS. The common TSS directs expression in brain, retina and testis and lacks both CRX and NRL binding (UCSC EST database[51]). Because the two transcripts differ in their first exons we were able to confirm by RT-PCR that the transcript from the rod TSS is upregulated ~200 times in postnatal retina, while the transcript from common TSS shows a smaller, though still significant increase (Fig 5F). We concluded that, in retina, rod photoreceptor transcription factors CRX and in less extent NRL are associated with rod TSS for genes with multiple TSS.

Tnfaip3 has been identified as a negative regulator of the NF- κ B pathway and has not previously been implicated in retina development or maintenance. One TSS of this gene demonstrates features of expression in rods namely CRX and NRL binding, as well as H3K4me2 accumulation later in development (Fig 5G). For the rod TSS we confirmed by RT-PCR that its transcript is upregulated over a thousand fold in postnatal retina, while the transcript from common TSS had a much smaller change in expression between PN1 and adult. (Fig 5H).

Rod TSSs are associated with DNase1 hypersensitive sites in retina

A catalog of retinal DNase1 Hypersensitive Sites (DHS) (University of Washington data) is available for 3 developmental stages, PN1 (1D), PN7 (1W), PN56 (8W) [35]. To compare rod-specific epigenetic signatures with DHS we monitored developmental changes in DHS for all 107 genes, estimating the stage-specific DHS occurrence in +/-1000bp window of each TSS. We calculated a ratio of DHS change during development by dividing DHS occupancy at PN56 with that at PN1 (PN56/PN1), and classified the TSS into three groups: DHS is down regulated during development, DHS is up regulated during development, or DHS is strongly up regulated during development. For the 72 rod genes with a single TSS, 94% showed a developmental increase in DHS; with 5 genes have DHS upregulated and 63 have DHS strongly upregulated (Fig 6A, upper panel). For the 41 rod TSS of the 35 genes with multiple TSS, 93% showed a DHS increase similar to the single TSS rod genes; with 3 genes have DHS upregulated and 35 have DHS strongly upregulated (Fig 6A, middle panel). For the 37 common TSS of these genes, 27 (73%) demonstrated a decrease in DHS during development (Fig 6A, lower right panel). Plotting individual changes in DHS for genes with multiple TSS (Fig 6B), indicated that the PN56 to PN1 ratios of the rod TSS showed a wide variation but on average were 5 times greater than the ratios for the common TSS. Control groups of genes showed no significant change in DHS during development, similarly to common TSS.

Genome-wide tracks for one gene, *Tmem229b* (Fig 6C), illustrates the developmental increase in chromatin accessibility at a DHS for a rod TSS that lines up with CRX binding. The *Tmem229b* gene is expressed in a number of tissues and its common TSS is associated with unique ESTs from whole brain and bone marrow, while the rod TSS gives rise to only retina ESTs (UCSC EST database). For the proposed isoform generated from the rod TSS we confirmed by RT-PCR that it is upregulated a thousand fold between PN1 and PN15, while the transcript from the common TSS had constant high level of expression (Fig 6D).

To confirm tissue specificity of DHS changes for the TSSs of genes with multiple TSS we compared temporal changes in the DHS around rod and common TSS (TSS +/- 1000bp) for

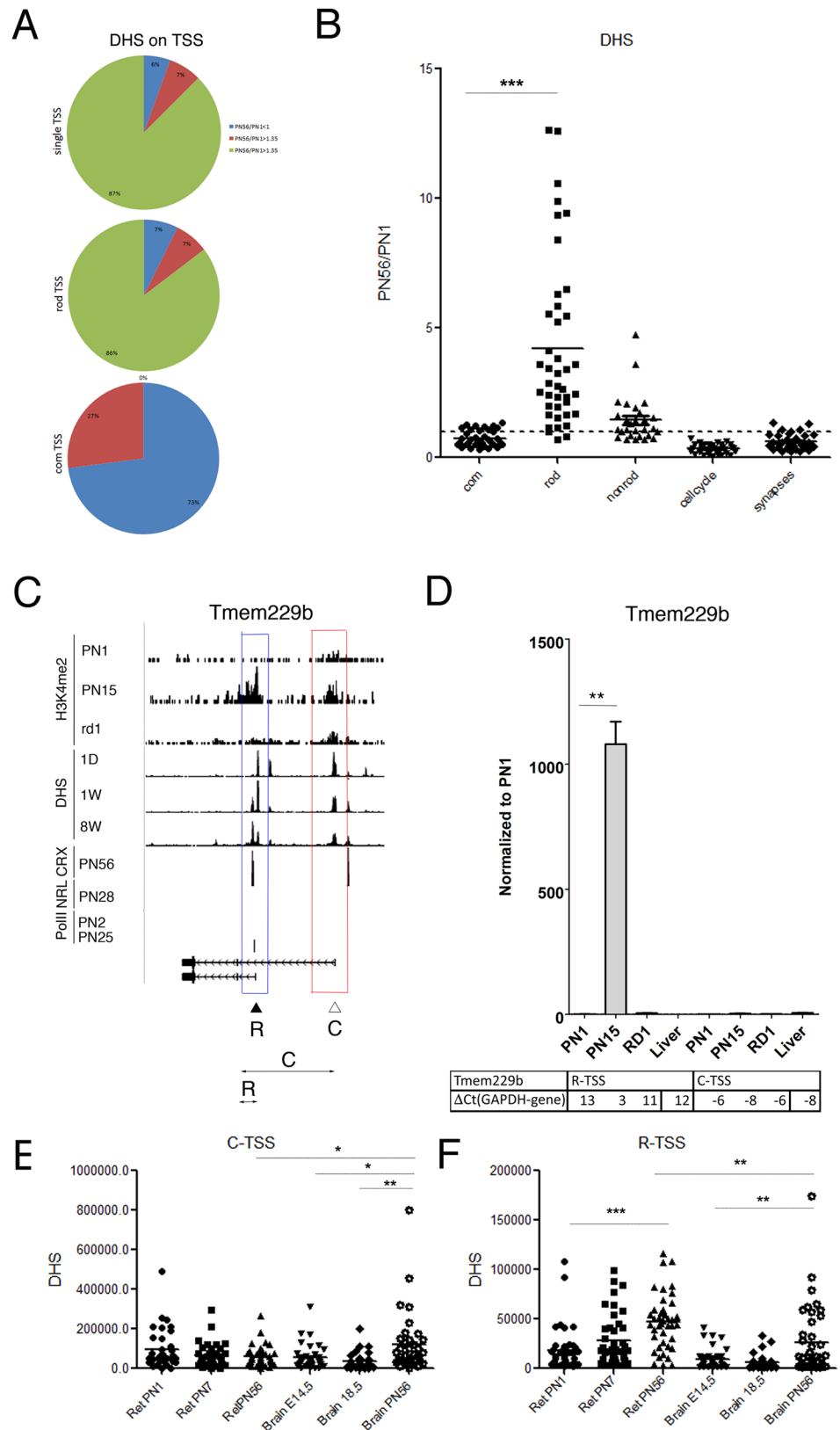


Fig 6. Rod TSSs are associated with DNase1 hypersensitive sites in retina. A. Number of TSS with different ratio of DHS changes during retina development (comparison of DHS occupancy at PN56 and PN1;

PN56/PN1) for rod genes with single TSS (upper panel) and for rod (middle panel) or common (bottom panel) TSS of retina genes with multiple TSS. Blue—no changes; red—low PN56/PN1; green—high PN56/PN1. **B.** Developmental changes in accessibility of different TSS by DHS (PN56/PN1 DHS at TSS \pm 1000bp) for rod and common TSS of retina genes with multiple TSS, compared with control groups of genes (non-rod, cell-cycle and synapse). ***— $p < 0.0001$. **C.** Combine genome-wide tracks of chromatin features, as in **Fig 1B** for *Tmem229b* gene. Common TSS is depicted as C/ red box, rod TSS—as R/ blue box. Position of the specific primer sets and PCR product that were used to assess and distinguish TSS-specific gene expression by RT-PCR depicted as double arrow below gene map. **D.** Relative gene expressions from rod and common TSS by RT-PCR with primer pairs depicted at **C** for *Tmem229b* for mouse retina samples at PN1, PN15, adult and RD1 mutant, compare with mouse liver. For comparison, normalized to *Gapdh* Δ Ct values for each sample are in table below. Experiments done in duplicates with three technical replicas; **— $p < 0.001$. **E, F.** Accessibility of different TSS by DHS (number of DHS at TSS \pm 1000bp) during mouse retina (PN1, PN7, PN56) and brain (E14.5, E18.5, PN56) developments for rod (**F**) and common TSS (**E**) of retina genes with multiple TSS.*— $p = 0.026$ (**E**); ***— $p = 0.0003$ (**F**); *— $p = 0.037/0.01$.

<https://doi.org/10.1371/journal.pone.0179230.g006>

retina (ENCODE PN1, PN7, PN56) and brain (ENCODE E14.5, E18.5, PN56) development. For the common TSS only DHS in the brain development group showed changes (**Fig 6E**). For the rod TSS there was a clear increase in DHS over time in retina but much smaller change in brain (**Fig 6F**).

From this we concluded that developmental changes in DHS at TSSs are consistent with their epigenetic profiles and can help indicate tissue-specific use of TSSs of ubiquitously expressed genes.

Rod TSSs are associated with PolII-binding sites in retina

As a further marker of transcriptional activity we examined PolII binding and estimated stage-specific PolII occupancy in \pm 1000bp window of each TSS. We calculated the PolII-binding changes during development by subtracting PolII occupancy at PN2 from PolII occupancy at PN25 (Δ PolII). We then clustered these developmental changes in PolII binding into three groups: no (or decrease) PolII binding, PolII binding increased slightly during development, and PolII increased substantially during development.

For the 72 rod genes with a single TSS, 67% showed a developmental increase in PolII binding; with 22 genes have slight and 26 have substantial increase of PolII binding (**Fig 7A**, upper panel). For the 41 rod TSS of the 35 genes with multiple TSS, 68% also showed a PolII binding increase similar to the single TSS rod genes; with 19 genes have slight and 9 have substantial increase of PolII binding (**Fig 7A**, middle panel). For the 37 common TSS, 28 (76%) have no PolII binding during development (**Fig 7A**, bottom panel).

We plotted the Δ PolII for each of the TSS for genes with multiple TSS and the control groups of genes (**Fig 7B**). Comparison between rod and common TSS shows 3.8 times an average increase in PolII-binding in vicinity of rod TSS (**Fig 7B**). The groups of control genes showed little change or even decreased in PolII-binding.

These changes are illustrated in the genome-wide tracks for *Kdm4c*, a histone demethylase implicated in development [52] (**Fig 7C**). PolII binding was detected at the rod TSS only in the adult whereas binding to the common TSS was the same at PN2 and PN25. As with many of the other genes in this class, the rod TSS had increasing levels of H3K4me2 with age and a DHS that increased in adult, while both were unchanged during development at the common TSS. Interestingly, CRX has binding sites at both *Kdm4c* TSSs, but NRL binds only to the rod TSS. We confirmed by RT-PCR that the proposed rod TSS generates a transcript that is up-regulated ~50 fold in postnatal retina, while the transcript from the common TSS has a constant high expression level (**Fig 7D**).

We conclude that developmentally regulated PolII-binding correlates with the expression from rod TSS, while common TSS mostly shows no developmental changes in PolII binding.

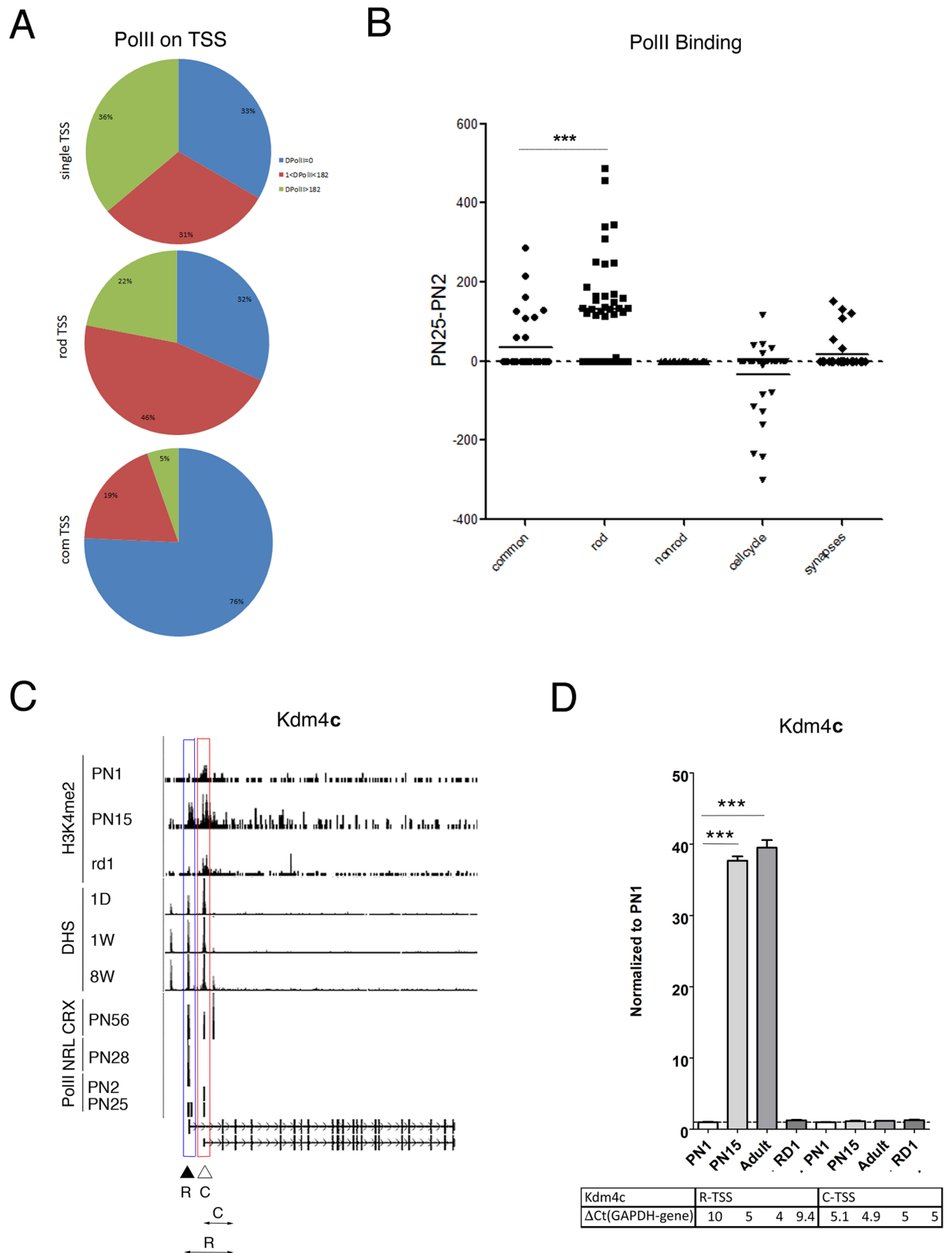


Fig 7. Rod TSSs are associated with PoII-binding sites in retina. **A.** Number of TSS with different ratio of PoII-binding during development (comparison of PoII occupancy at PN25 and PN2; PN25-PN2) for rod-specific genes with single TSS (upper panel) and

for rod (middle panel) or common (bottom panel) TSS of retina genes with multiple TSS. Blue—no changes; red—low PN25-PN2; green—high PN25-PN2. **B.** Developmental changes in PolII-binding (PN25-PN2 at TSS \pm 1000bp) for rod and common TSS of retina genes with multiple TSS, compared with PolII-binding at TSS for control groups of genes (non-rod, cell-cycle and synapse). ***— $p < 0.0001$. **C.** Combined genome-wide tracks of chromatin features, as in Fig 1B for *Kdm4c* gene. Common TSS is depicted as C/ red box, rod TSS—as R/ blue box. Position of the specific primer sets and PCR product that were used to assess and distinguish TSS-specific gene expression by RT-PCR depicted as double arrow below gene map. **D.** Relative gene expressions from rod and common TSS by RT-PCR with primer pairs depicted at C for *Kdm4c* for mouse retina samples at PN1, PN15, adult and RD1 mutant. For comparison, normalized to *Gapdh* Δ Ct values for each sample are in table below. Experiments done in duplicates with three technical replicas; ***— $p < 0.0001$.

<https://doi.org/10.1371/journal.pone.0179230.g007>

Using chromatin features to predict tissue specific new TSS

To examine the predictive power of our approach we compared three clusters: TSS for known rod genes, common TSS and newly predicted rod TSS. Each TSS was characterized by 13 chromatin features (as on Fig 1B), such as H3K4me2 accumulation at E17, PN1, PN7, PN15 and in RD1; DNase hypersensitive sites at 1D, 1W, W8; binding of PolII at PN2 and PN25; binding of CRX (samples S1 and S2) and NRL. To evaluate our clusters we used the Silhouette coefficient (SC), which combines the concepts of cluster cohesion and cluster separation. An average SC greater than 0.5 indicates reasonable partitioning of data while average SC less than 0.2 means that data do not exhibit cluster structure. Comparison between clusters of TSS for known rod genes with common TSS gave an average SC = 0.75; and comparison between clusters of newly predicted rod TSS and common TSS gave an average SC = 0.5 (Fig 8A). These data demonstrate that common TSS is a reasonable cluster and it is separated from known and newly predicted rod TSS clusters. Comparison between known and newly predicted rod TSS clusters gave an average SC = -0.02, indicating that no separation exists between these two clusters or that all these TSS belong to the same cluster, rod TSS. Thus we have been able to recognize rod TSS with a high degree of certainty.

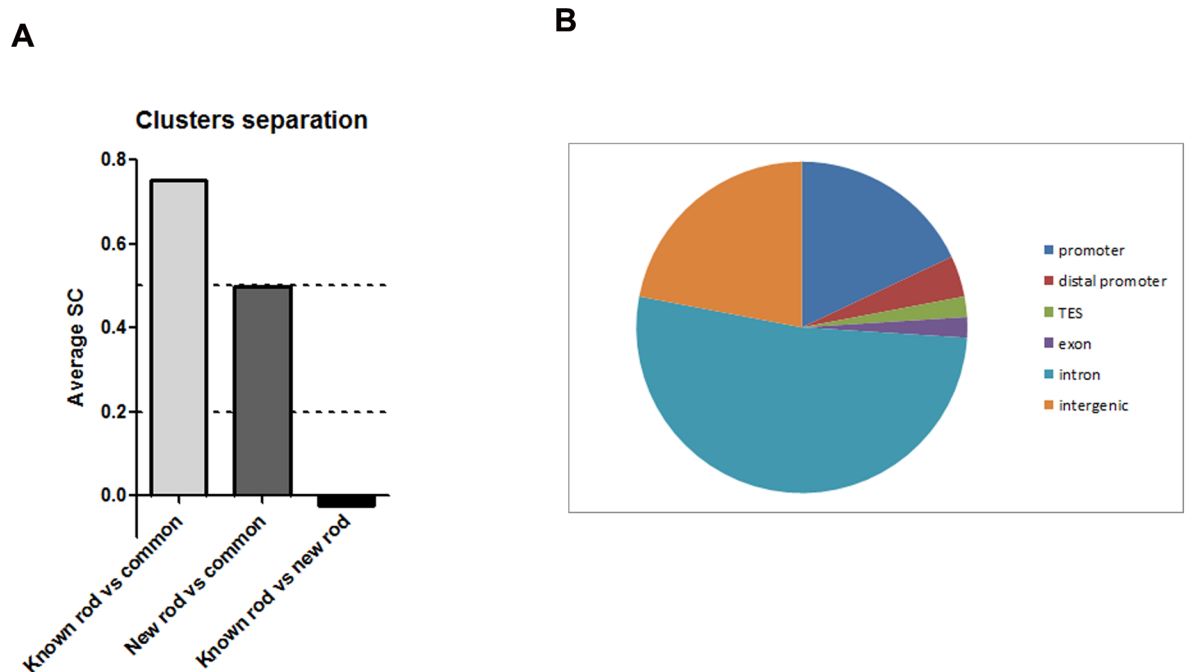


Fig 8. Predictive power of epigenetic profiles. A. Cluster separation for different TSS groups. SC—Silhouette coefficient. B. Pie chart representation of location in the genome sites that have chromatin signature characteristic for rod TSS.

<https://doi.org/10.1371/journal.pone.0179230.g008>

We then tested whether our approach also had predictive power and could identify new rod TSS that were not in the current TSS database for RefSeq. Our preliminary data indicated that there are around 600 sites on mouse genome that exhibit a chromatin signature characteristic for rod TSS and 12% of these sites are located in gene loci whose human homologs are implicated in human retinal diseases in RetNet. 22% of these 600 sites are located in promoters, 2% are in exons, 2% are in Transcription End Site (TES), 22% are in the intergenic areas of the genome and 52% are in the introns (Fig 8B). Intergenic sites could represent such chromatin regulatory sites, as enhancers or insulators, while sites located in introns could be potential new rod TSS.

On mouse chromosomes 1 and 2 we found 80 loci with rod TSS features, we then compared expression from this new predicted rod TSS with expression from common known TSS for four of these loci. For three out of four loci we have confirmed rod specificity for newly predicted sites. Fig 9 presents examples of the prediction of new rod TSS.

Cacnb2, encoding one of the beta subunits of Ca⁺⁺ voltage-gated channels, is widely expressed in different tissues [53, 54]. Based on chromatin features around the gene locus we predicted that *Cacnb2* has an additional rod TSS (Fig 9A). We designed 2 pairs of PCR primers to test expression from the newly predicted rod and the common TSS. Transcription from the rod TSS is highly upregulated in adult tissue and is less in liver or RD1 retina, while transcription from common TSS is not dependent on age and is slightly upregulated in RD1 retina in to the level in liver (Fig 9B).

Ndufs2 encodes one subunit of NADH: ubiquinone oxidoreductase (complex I) of the mitochondrial respiratory chain [55]. The known TSS of this gene in retina has the epigenetic profile of constantly expressed gene: DHS and H3K4me2 accumulation are constant during retina development, and PolII binding is similar at PN2 and PN25 (Fig 9C). Between exon 8 and 9 we found a chromatin region highly resembling the TSS profile of a rod gene: CRX and NRL binding, PolII binding only at PN25 and DHS and H3K4me2 increased at the end of retina development (Fig 9C). We probed expression from the common TSS and this new rod TSS in cDNA from retina of different ages and compared this with cDNA from liver and RD1 mutant retina. Expression from the rod TSS was upregulated during retina maturation, but was much lower in RD1 retina and liver, while transcription from common TSS showed little change in retinal development and was higher in liver (Fig 9D).

Map2 encodes a microtubule-associated protein and produces multiple transcripts by alternative splicing that give rise to several developmentally regulated isoforms of the Map2 protein [56]. The known TSSs of this gene in retina have the epigenetic profile of a constantly expressed gene (Fig 9E). After exon 4 we found chromatin features highly resembling the TSS profile for rod genes (Fig 9E). We probed expression from common TSS and this new rod TSS in cDNA from retinas of different ages and compared this with cDNA from liver and retina from RD1 mutant retina. Expression from the rod TSS was upregulated during retina maturation, and present only at much lower levels in RD1 retina and hardly at all in liver, while transcription from the common TSS was maintained at high levels through retinal development (Fig 9F).

Many of the genes that have a rod TSS and a common TSS are also expressed in other tissues. To explore TSS usage in variety of tissue we used accumulation of H3K4me3 as a marker of active TSS for several of these genes. We choose four new rod genes exhibiting both common and rod TSS. As shown in Fig 10, *Kdm4c*, *Tnfaip3*, *Tmem229b* and *Ablim1* show active transcription in many, but not all, tissues (Fig 10A–10D). The use of extra TSS varied considerably. For *Kdm4c* we detected a rod TSS but no additional TSS in any of the other tissue. *Tnfaip3* showed an additional TSS in small intestine that yields an even smaller transcript than the rod specific form. *Tmem229* showed 2 types of alternative TSS. One was restricted to

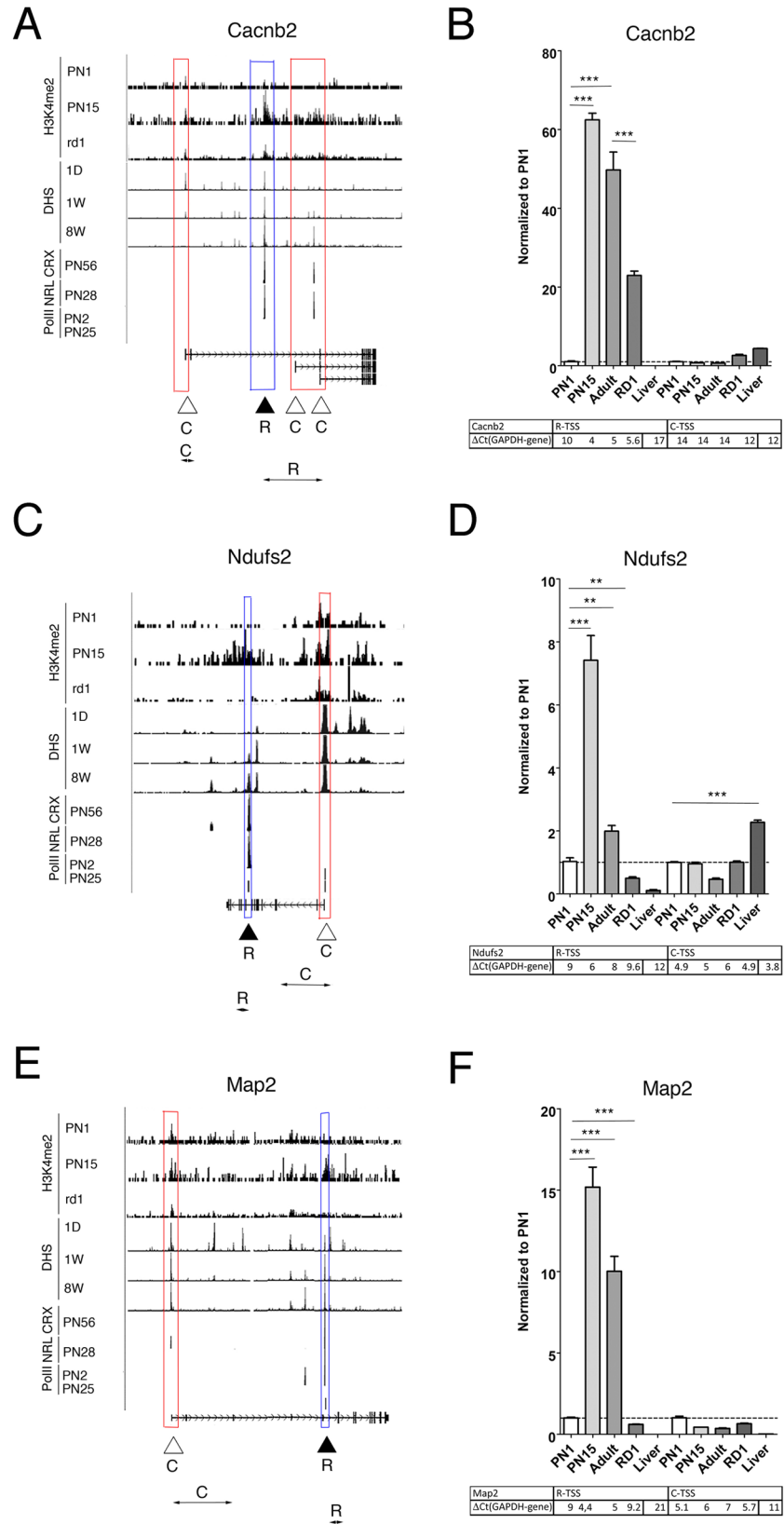


Fig 9. Chromatin features predict tissue specific new TSS. A, C, E. Combined genome-wide tracks of chromatin features, as in Fig 1B for *Cacnb2* (A), *Ndufs2* (C) and *Map2* (E) genes. Common TSS is depicted as C/ red box, predicted rod TSS—as R/ blue box. Position of the specific primer sets and PCR product that were used to assess and distinguish TSS-specific gene expression by RT-PCR depicted as double arrow below gene maps. **B, D, F.** Relative gene expressions from predicted rod and common TSS by RT-PCR with primer pairs depicted at A, C, E for *Cacnb2* (B), *Ndufs2* (D) and *Map2* (F) for mouse retina samples at PN1, PN15, adult and RD1 mutant, compare with mouse liver. For comparison, normalized to *Gapdh* Δ Ct values for each sample are in table below. Experiments done in duplicates with three technical replicas; ***— $p < 0.0001$, **— $p < 0.001$.

<https://doi.org/10.1371/journal.pone.0179230.g009>

cerebellum and other was used in multiple tissues. *Ablim1* has multiple TSS. In addition to the rod TSS, two of the other TSS are used in multiple tissues (and were labeled common TSS in Fig 3). We also detected three additional potential TSS used by one or few tissues.

The mouse retina provides a good, though imperfect, model system to aid understanding of the human retina. As proof of concept, that rod TSS are conserved between human and mouse and that our approach to find tissue-preferential TSS is applicable to other organisms, we compared our data with RNA-seq data from human retina [57], which is available on UCSC Human Genome Browser. From our list of 107 genes—85% have same or very similar gene locus structure in mouse and human, and 81% of TSS in human demonstrate patterns of RNA-seq corresponding to the epigenetic profile in the mouse genome.

We found, that if the structure of the mouse gene locus is similar to that of the human, rod TSS could be easily distinguished from common TSS (Fig 11). For example, the *TNFAIP3* gene is highly expressed in human retina from the same ATSS as in mouse (Fig 11A and Fig 5G), the same is true for *PLA2G5*, a gene from the RetNet list for mapped loci and genes causing retinal diseases that has been implicated in recessive benign fleck retina [58] (Fig 11B). RefSeq data shows that *CACNB2* gene in mouse as in human has four alternative TSS. Epigenetic profiles in mouse retina pointed out an intron in the middle of the gene with increase in H3K4me2 accumulation and DHS in adult and CRX- and NRL-binding. We predicted, and confirmed, that *Cacnb2* has an additional rod TSS in the mouse (Fig 9A and 9B) that had previously been defined as an intron. The homologous region in human is also defined as an intron in RefSeq. The human retina RNA-seq data demonstrates an accumulation of RNA reads at this site, suggesting that in human too there is an exon produced from a novel rod TSS (Figs 11C and 9B).

All of these examples indicate that there is a rich landscape of alternative TSS yet to be explored and that the methods we have outlined can provide ways to identify them.

Discussion

In addition to generating different transcripts, the use of alternative transcription start sites allows separate spatial and temporal regulation of transcription in different cell types and, when multiple TSSs are used in a single cell, allows separate control of constitutive and regulated levels of expression. In the present study, we have explored the properties and use of alternative TSS in the mammalian retina and specifically in one cell type, the rod photoreceptor. In the adult rodent retina, rods comprise at least 85% of the cells, making it easy to measure rod gene properties against a small background from other cell types. We have used the term rod TSS and rod gene in this work. We are confident that these transcripts are expressed in rods, based on the temporal expression patterns and significant loss of expression ($p < 0.0001$), in the RD1 mutant retina [31]. We cannot, however, exclude the possibility that some of these are also expressed in some other retinal cells, or in other tissues. The results shown in Fig 10 indicate that tissue restricted TSS may be used by a variable number of tissues or cell types, but not by all.

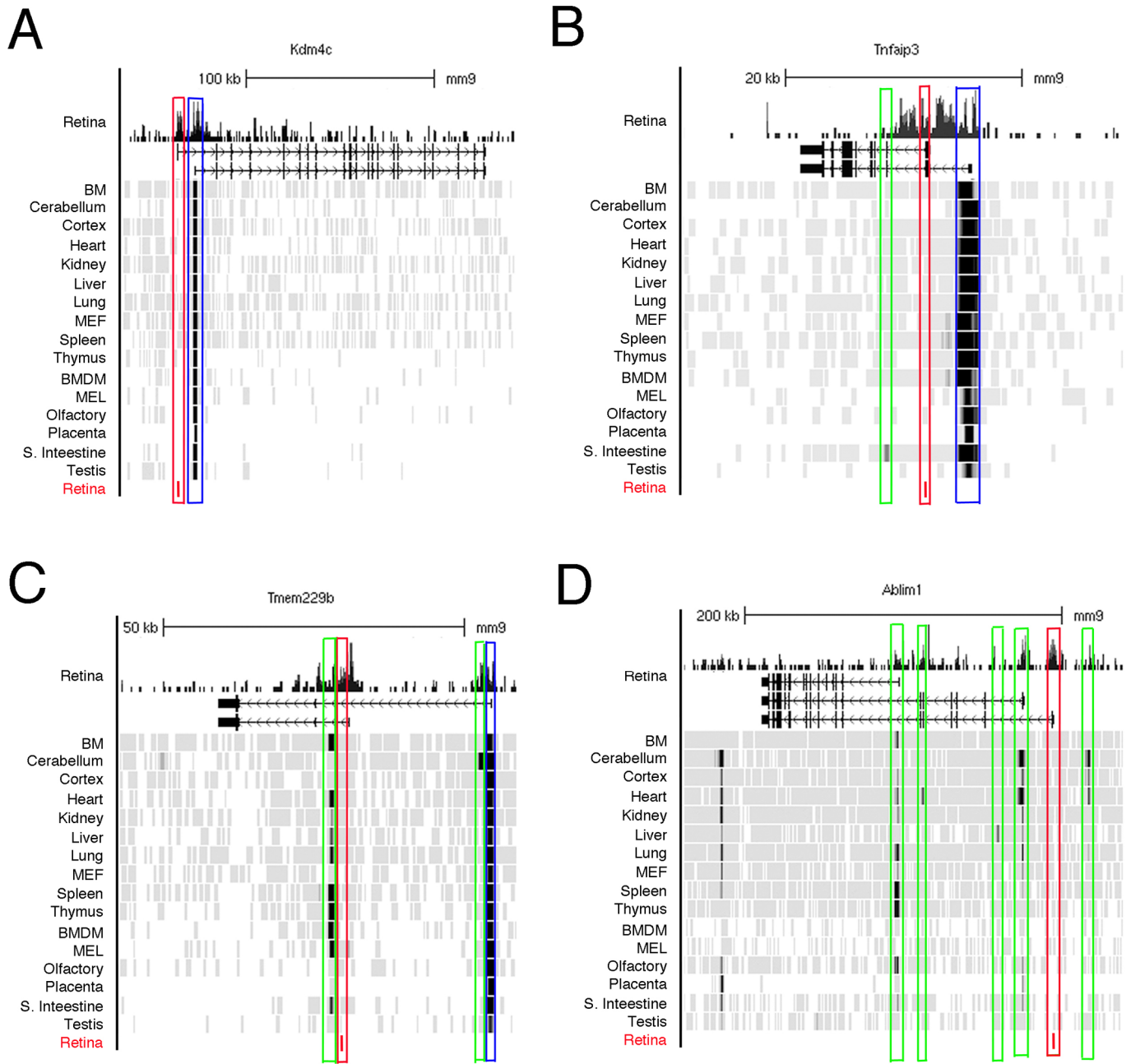


Fig 10. Use of alternative TSS for tissue specificity is universal. A-D. Combined genome-wide tracks of H3K4me3 accumulation (UCSC tracks for mouse browser mm9) for different mouse tissue and cells (bone marrow, cerebellum, cortex, heart, kidney, liver, lung, mouse embryonic fibroblasts, spleen, thymus, bone marrow derived macrophages, murine erythroleukemia cells, olfactory bulb, placenta, small intestine and testis) for the new rod genes: *Kdm4c* (A) (also see Fig 7C for this gene); *Tnfaip3* (B) (also see Fig 5H for this gene); *Tmem229b* (C) (also see Fig 6D for this gene); *Ablim1* (D) (also see Fig 3A for this gene). In each case common TSS is depicted as a blue box, rod TSS—as a red box and other tissue-specific TSS—as green boxes. For reference on the top of each figure panel presents the genome-wide track of H3K4me2 for retina at PN15 and gene position and locus structure.

<https://doi.org/10.1371/journal.pone.0179230.g010>

We previously showed that all previously known rod genes (a total of 36) shared a specific developmental epigenetic signature, namely a de novo accumulation of H3K4me2 and a lack of H3K27me3 around the promoter, while some key precursors genes were losing H3K4me2

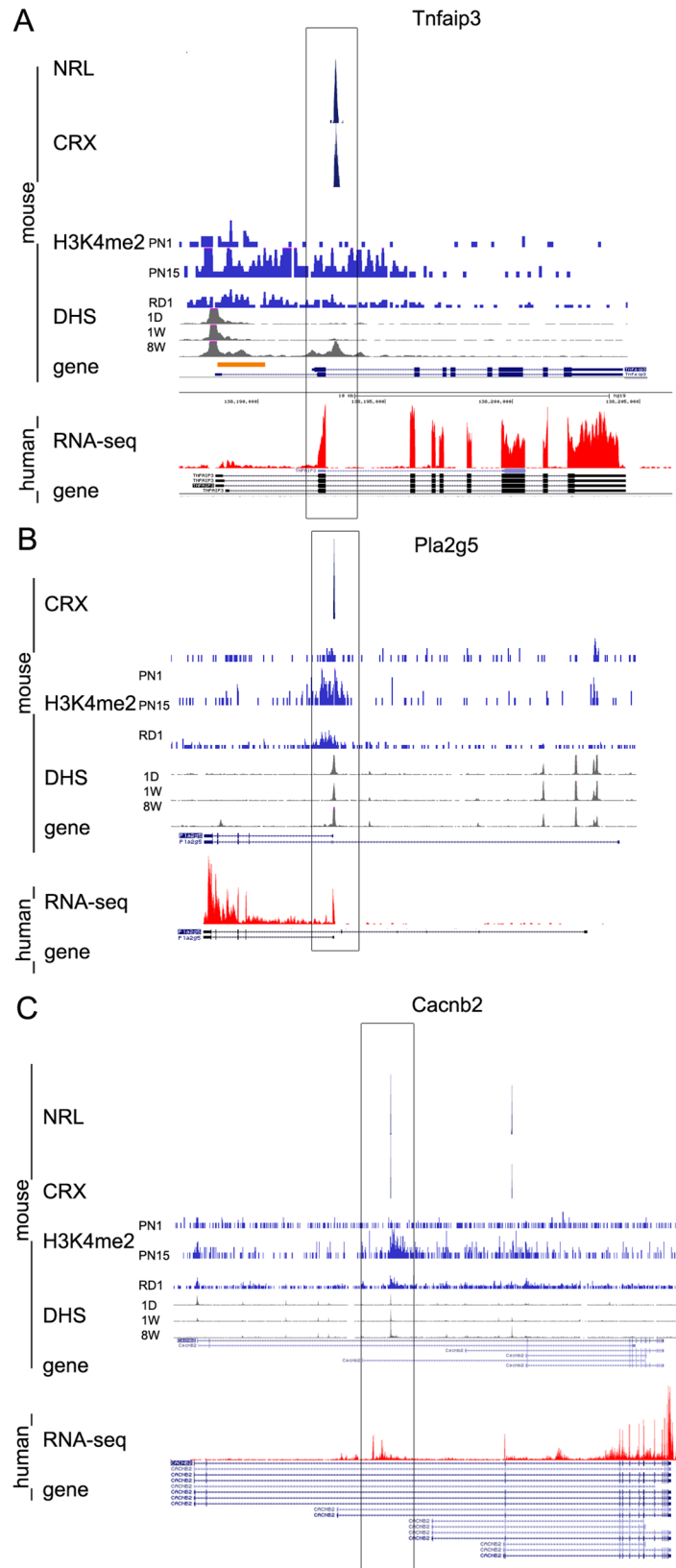


Fig 11. Rod TSS are conserved between human and mouse. A, B, C. Combined genome-wide tracks of chromatin features in mouse retina (developmental changes in H3K4me2 occupancy (PN1, PN15 and RD1 PN30), CRX-binding (PN56), NRL-binding (PN28), developmental changes in DHS (1D, 1W, 8W)) and RNA-seq for human retina for genes TNFAIP3 (A), PLA2G5 (B) and CACNB2 (C) genes. Predicted rod TSS that shows conservation between mouse and human depicted as a black box.

<https://doi.org/10.1371/journal.pone.0179230.g011>

accumulation during development [31]. Inhibiting a key histone demethylase, LSD1, led to maintained levels of H3K4me2, sustained expression of precursor genes and a block in the expression of rod specific genes, confirming the active role of epigenetic modification in guiding developmental gene expression [59]. We used our rod epigenetic signature to screen the whole genome and found an additional 71 genes that were candidates for expression in rods.

We have now examined the TSSs of these genes and tested a group of chromatin and epigenetic features to explore and compare whether these characteristics can determine cell type specific TSSs, particularly in those genes with multiple TSSs. The features that we used were binding of the characteristic rod-specific transcription factors CRX [33] and NRL [32], developmental changes of DHS at the TSS [35], PolII binding [34] together with our developmental profiles of H3K4me2 accumulation.

CRX and NRL are transcription factors that have been well characterized as regulators of expression of rod specific genes involved in visual transduction. The 36 known rod genes, with the exception of *Cacna1f*, had CRX binding at one TSS. This association of CRX was also seen with the new rod genes, such that 76% of the genes with a single TSS and 73% of the rod TSS in genes with multiple TSS had CRX binding (Fig 5A). This is a strong correlation and emphasizes the importance of CRX in rod differentiation. It will be interesting to determine whether the other approximately 25% of genes use transcription factors are carrying out the same role as CRX.

NRL is also known to regulate the transcription of rod-specific genes but some known rod genes do not have measurable NRL binding. Overall, we found that 61% of the 72 rod genes with a single TSS showed NRL binding as did 56% of the rod TSS in genes with multiple TSS. This indicates that NRL is responsible for regulating the transcription of only a subset of rod genes and that other transcription factors, possibly other basic motif-leucine zipper family members, regulate the others.

DHS is a marker of open chromatin sites and, when present at TSS, is an excellent indicator of active transcription from that site. All known rod genes show a developmental increase in DHS that correlates with increased gene expression. A similar change in DHS was found for the new rod genes that had only one TSS. For those genes with multiple TSS it was clear that one behaved like a rod TSS and the other was active throughout development. This would suggest that DHS level can be a useful diagnostic of tissue-specific TSS when different ages can be compared.

In general, PolII binding can also help define rod TSS, but is less effective than the other markers. A greater proportion of genes in both the known rod and the new rod classes had no change in PolII binding between PN2 and PN25. The data set for PolII binding that available for retina and have been used here is based on ChIP-on-Chip assay and because of this have limited genomic coverage. Future studies based on ChIP-Seq methods and carried out at several developmental points could greatly improve our knowledge of PolII binding. Additionally, it is possible that PN2 is close to the time when many rod genes are beginning to assemble transcription complexes and so PolII is already bound, thus lowering the change. Determining PolII binding at an earlier time point, such as E17 might give better discrimination.

In many cases the TSS we have defined as “common” are active at high levels throughout development. One question that arises is whether they are used only by cells other than rods or whether they generate transcripts in rods as well as other cell types. Since the transcriptional

level from common TSS in RD1 mice is comparable with the transcriptional level from common TSS in wild type mice (average $\Delta C_{t_{rd1}}$ (Gapdh-gene) = 5.2; average $\Delta C_{t_{wt}}$ (Gapdh-gene) = 5.1) it is likely that the common TSS are active in all retina cells and that the rod TSS serves to add additional transcripts to meet the high protein synthesis demand of these very active cells. Where the multiple TSS generate different proteins, this may provide a way of having multiple ways of regulating protein activity within a single cell type.

An alternative to our epigenetic approach to find new TSS is RNA-seq. Several RNA-seq data sets are available now for mouse retina [60–63], but in some instances comparison between these data sets show dissimilarities, possibly because of different genetic backgrounds or different methods for cell purification [62]. The power of our methodology is in comparing temporal epigenomic profiles at several developmental stages and combination of epigenetic signatures and TF binding data that can be used to define TSSs that are currently not recognized. In our study of 35 genes with multiple TSS, (6 are known rod-specific genes, whereas the other 29 identify new genes), we have tested and compared transcription for common and rod TSS at PN1 and PN15/adult retina for 11 of them. 9 showed the expected patterns (6—shown in Figs 3, 5, 6 and 7; 3—data not shown), so it gives us an 18% false positive rate. The false negative rate is very small as from 98 control genes (Figs 2D; 5B; 6B) none had a rod pattern on their TSS. From our preliminary analysis of sites along mouse genome we suggest that many such TSS remain to be discovered. We tested the use of the new predicted TSS in three such examples (Fig 9). Each had a previously defined TSS (“common” by our criteria) and, using our analysis, each also has a rod TSS. In each case, we were able to verify a developmentally regulated transcript from this new TSS in retina that was lost in RD1 mice.

We found >80% concordance of mouse and human rod TSS, with biological confirmation of the human predictions from available RNA-seq data. This indicates that understanding the regulation of gene expression in the mouse will provide valuable information about the normal and diseased human retina, and also a good model to study therapeutic interventions that seek to modulate the expression of disease genes.

We have applied the approach of using a group of chromatin and epigenetic features to identify transcriptional complexity in one cell type, the rod photoreceptor. This approach gives the most confident results about transcription tissue specificity when chromatin markers at a particular TSS can be compared in different tissues (Fig 10) or at several time points in development, as we have shown here for H3K4me2 and DHS.

Based on our initial studies, we would predict that a similar approach can be used in other tissues and cell types to refine the current maps of TSS. While genes uniquely associated with the function of a cell, such as visual transduction pathway genes in rods, frequently have a single TSS, other genes with multiple TSS may represent the evolution of additional controls of the levels and nature of transcripts to better regulate cell-specific metabolism and function.

Supporting information

S1 File. Combined supporting figures. This file contains two figures labeled A, B. (PDF)

S1 Table. PCR primers. (XLSX)

S2 Table. List of the 107 rod genes with single and multiple TSS. (XLSX)

S3 Table. Epigenetic characteristic of rod genes TSS. (XLSX)

S4 Table. List of genes for cluster 1 and 2.

(XLSX)

S5 Table. List of three groups of control genes.

(XLSX)

Author Contributions**Conceptualization:** CJB SSMZ.**Data curation:** ACS.**Formal analysis:** ACS EYP.**Funding acquisition:** CJB.**Investigation:** EYP CY.**Methodology:** EYP ACS SSMZ.**Project administration:** CJB SSMZ EYP.**Software:** ACS.**Supervision:** CJB.**Validation:** EYP.**Visualization:** EYP ACS SSMZ.**Writing – original draft:** CJB EYP.**Writing – review & editing:** EYP SSMZ CJB.**References**

1. Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, et al. Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res.* 2011; 21(8):1260–72. <https://doi.org/10.1101/gr.120535.111> PMID: 21712398; PubMed Central PMCID: PMC3149493.
2. Shabalina SA, Ogurtsov AY, Spiridonov NA, Koonin EV. Evolution at protein ends: major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals. *Nucleic Acids Res.* 2014; 42(11):7132–44. <https://doi.org/10.1093/nar/gku342> PMID: 24792168; PubMed Central PMCID: PMC4066770.
3. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell.* 2016; 166(5):1147–62.e15. <https://doi.org/10.1016/j.cell.2016.07.025> PMID: 27565344.
4. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet.* 2008; 24(4):167–77. <https://doi.org/10.1016/j.tig.2008.01.008> PMID: 18329129.
5. Sun H, Wu J, Wickramasinghe P, Pal S, Gupta R, Bhattacharyya A, et al. Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids Res.* 2011; 39(1):190–201. <https://doi.org/10.1093/nar/gkq775> PMID: 20843783; PubMed Central PMCID: PMC3017616.
6. Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, et al. A promoter-level mammalian expression atlas. *Nature.* 2014; 507(7493):462–70. <https://doi.org/10.1038/nature13182> PMID: 24670764; PubMed Central PMCID: PMC4529748.
7. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science.* 2016; 352(6292):1413–6. <https://doi.org/10.1126/science.aad9868> PMID: 27313038.

8. Wang X, Hou J, Quedenau C, Chen W. Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol Syst Biol.* 2016; 12(7):875. Epub 2016/07/18. PMID: [27430939](https://pubmed.ncbi.nlm.nih.gov/27430939/); PubMed Central PMCID: [PMCPMC4965872](https://pubmed.ncbi.nlm.nih.gov/PMC4965872/). <https://doi.org/10.15252/msb.20166941>
9. Tamarkin-Ben-Harush A, Vasseur JJ, Debart F, Ulitsky I, Dikstein R. Cap-proximal nucleotides via differential eIF4E binding and alternative promoter usage mediate translational response to energy stress. *Elife.* 2017; 6. Epub 2017/02/08. <https://doi.org/10.7554/eLife.21907> PMID: [28177284](https://pubmed.ncbi.nlm.nih.gov/28177284/); PubMed Central PMCID: [PMCPMC5308895](https://pubmed.ncbi.nlm.nih.gov/PMC5308895/).
10. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 2007; 448(7153):553–60. PMID: [17603471](https://pubmed.ncbi.nlm.nih.gov/17603471/). <https://doi.org/10.1038/nature06008>
11. Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, Zhu J, et al. Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.* 2011; 7(1):e1001274. <https://doi.org/10.1371/journal.pgen.1001274> PMID: [21249180](https://pubmed.ncbi.nlm.nih.gov/21249180/); PubMed Central PMCID: [PMCPMC3020932](https://pubmed.ncbi.nlm.nih.gov/PMC3020932/).
12. Mortazavi A, Pepke S, Jansen C, Marinov GK, Ernst J, Kellis M, et al. Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps. *Genome Res.* 2013; 23(12):2136–48. <https://doi.org/10.1101/gr.158261.113> PMID: [24170599](https://pubmed.ncbi.nlm.nih.gov/24170599/); PubMed Central PMCID: [PMCPMC3847782](https://pubmed.ncbi.nlm.nih.gov/PMC3847782/).
13. Natarajan A, Yardimci GG, Sheffield NC, Crawford GE, Ohler U. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res.* 2012; 22(9):1711–22. <https://doi.org/10.1101/gr.135129.111> PMID: [22955983](https://pubmed.ncbi.nlm.nih.gov/22955983/); PubMed Central PMCID: [PMCPMC3431488](https://pubmed.ncbi.nlm.nih.gov/PMC3431488/).
14. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* 2012; 13(9):R53. <https://doi.org/10.1186/gb-2012-13-9-r53> PMID: [22950368](https://pubmed.ncbi.nlm.nih.gov/22950368/); PubMed Central PMCID: [PMCPMC3491397](https://pubmed.ncbi.nlm.nih.gov/PMC3491397/).
15. Zhang SS, Xu X, Li J, Liu MG, Zhao H, Soares MB, et al. Comprehensive in silico functional specification of mouse retina transcripts. *BMC Genomics.* 2005; 6:40. <https://doi.org/10.1186/1471-2164-6-40> PMID: [15777472](https://pubmed.ncbi.nlm.nih.gov/15777472/); PubMed Central PMCID: [PMCPMC1083414](https://pubmed.ncbi.nlm.nih.gov/PMC1083414/).
16. Zhang SS, Xu X, Liu MG, Zhao H, Soares MB, Barnstable CJ, et al. A biphasic pattern of gene expression during mouse retina development. *BMC Dev Biol.* 2006; 6:48. <https://doi.org/10.1186/1471-213X-6-48> PMID: [17044933](https://pubmed.ncbi.nlm.nih.gov/17044933/); PubMed Central PMCID: [PMCPMC1633734](https://pubmed.ncbi.nlm.nih.gov/PMC1633734/).
17. Livesey FJ, Furukawa T, Steffen MA, Church GM, Cepko CL. Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene *Crx*. *Curr Biol.* 2000; 10(6):301–10. PMID: [10744971](https://pubmed.ncbi.nlm.nih.gov/10744971/).
18. Blackshaw S, Fraioli RE, Furukawa T, Cepko CL. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. *Cell.* 2001; 107(5):579–89. PMID: [11733058](https://pubmed.ncbi.nlm.nih.gov/11733058/).
19. Mu X, Zhao S, Pershad R, Hsieh TF, Scarpa A, Wang SW, et al. Gene expression in the developing mouse retina by EST sequencing and microarray analysis. *Nucleic Acids Res.* 2001; 29(24):4983–93. PMID: [11812828](https://pubmed.ncbi.nlm.nih.gov/11812828/); PubMed Central PMCID: [PMCPMC97568](https://pubmed.ncbi.nlm.nih.gov/PMC97568/).
20. Farjo R, Yu J, Othman MI, Yoshida S, Sheth S, Glaser T, et al. Mouse eye gene microarrays for investigating ocular development and disease. *Vision Res.* 2002; 42(4):463–70. PMID: [11853762](https://pubmed.ncbi.nlm.nih.gov/11853762/).
21. Sharon D, Blackshaw S, Cepko CL, Dryja TP. Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). *Proc Natl Acad Sci U S A.* 2002; 99(1):315–20. <https://doi.org/10.1073/pnas.012582799> PMID: [11756676](https://pubmed.ncbi.nlm.nih.gov/11756676/); PubMed Central PMCID: [PMCPMC117558](https://pubmed.ncbi.nlm.nih.gov/PMC117558/).
22. Dorrell MI, Aguilar E, Weber C, Friedlander M. Global gene expression analysis of the developing postnatal mouse retina. *Invest Ophthalmol Vis Sci.* 2004; 45(3):1009–19. PMID: [14985324](https://pubmed.ncbi.nlm.nih.gov/14985324/).
23. Siegert S, Cabuy E, Scherf BG, Kohler H, Panda S, Le YZ, et al. Transcriptional code and disease map for adult retinal cell types. *Nat Neurosci.* 2012; 15(3):487–95, S1-2. <https://doi.org/10.1038/nn.3032> PMID: [22267162](https://pubmed.ncbi.nlm.nih.gov/22267162/).
24. Brooks MJ, Rajasimha HK, Roger JE, Swaroop A. Next-generation sequencing facilitates quantitative analysis of wild-type and *Nrl(-/-)* retinal transcriptomes. *Mol Vis.* 2011; 17:3034–54. PMID: [22162623](https://pubmed.ncbi.nlm.nih.gov/22162623/); PubMed Central PMCID: [PMCPMC3233386](https://pubmed.ncbi.nlm.nih.gov/PMC3233386/).
25. Fullwood MJ, Wei CL, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* 2009; 19(4):521–32. <https://doi.org/10.1101/gr.074906.107> PMID: [19339662](https://pubmed.ncbi.nlm.nih.gov/19339662/); PubMed Central PMCID: [PMCPMC3807531](https://pubmed.ncbi.nlm.nih.gov/PMC3807531/).
26. Matsumoto K, Suzuki A, Wakaguri H, Sugano S, Suzuki Y. Construction of mate pair full-length cDNAs libraries and characterization of transcriptional start sites and termination sites. *Nucleic Acids Res.* 2014; 42(16):e125. <https://doi.org/10.1093/nar/gku600> PMID: [25034687](https://pubmed.ncbi.nlm.nih.gov/25034687/); PubMed Central PMCID: [PMCPMC4176323](https://pubmed.ncbi.nlm.nih.gov/PMC4176323/).

27. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals wide-spread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.* 2013; 23(1):169–80. <https://doi.org/10.1101/gr.139618.112> PMID: 22936248; PubMed Central PMCID: PMC3530677.
28. Yamashita R, Wakaguri H, Sugano S, Suzuki Y, Nakai K. DBTSS provides a tissue specific dynamic view of Transcription Start Sites. *Nucleic Acids Res.* 2010; 38(Database issue):D98–104. <https://doi.org/10.1093/nar/gkp1017> PMID: 19910371; PubMed Central PMCID: PMC2808897.
29. Suzuki A, Wakaguri H, Yamashita R, Kawano S, Tsuchihara K, Sugano S, et al. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.* 2015; 43(Database issue):D87–91. <https://doi.org/10.1093/nar/gku1080> PMID: 25378318; PubMed Central PMCID: PMC34383915.
30. Yasuda M, Tanaka Y, Nishiguchi KM, Ryu M, Tsuda S, Maruyama K, et al. Retinal transcriptome profiling at transcription start sites: a cap analysis of gene expression early after axonal injury. *BMC Genomics.* 2014; 15:982. <https://doi.org/10.1186/1471-2164-15-982> PMID: 25407019; PubMed Central PMCID: PMC34246558.
31. Popova EY, Xu X, DeWan AT, Salzberg AC, Berg A, Hoh J, et al. Stage and gene specific signatures defined by histones H3K4me2 and H3K27me3 accompany mammalian retina maturation in vivo. *PLOS One.* 2012; 7(10):e46867. Epub 2012/10/12. <https://doi.org/10.1371/journal.pone.0046867> PMID: 23056497; PubMed Central PMCID: PMC3467275.
32. Hao H, Kim DS, Klocke B, Johnson KR, Cui K, Gotoh N, et al. Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL targetome analysis. *PLoS Genet.* 2012; 8(4):e1002649. <https://doi.org/10.1371/journal.pgen.1002649> PMID: 22511886; PubMed Central PMCID: PMC3325202.
33. Corbo JC, Lawrence KA, Karlstetter M, Myers CA, Abdelaziz M, Dirkes W, et al. CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. *Genome Res.* 2010; 20(11):1512–25. <https://doi.org/10.1101/gr.109405.110> PMID: 20693478; PubMed Central PMCID: PMC2963815.
34. Tummala P, Mali RS, Guzman E, Zhang X, Mitton KP. Temporal ChIP-on-Chip of RNA-Polymerase-II to detect novel gene activation events during photoreceptor maturation. *Mol Vis.* 2010; 16:252–71. PMID: 20161818; PubMed Central PMCID: PMC2822553.
35. Wilken MS, Brzezinski JA, La Torre A, Siebenthal K, Thurman R, Sabo P, et al. DNase I hypersensitivity analysis of the mouse brain and retina identifies region-specific regulatory elements. *Epigenetics Chromatin.* 2015; 8:8. <https://doi.org/10.1186/1756-8935-8-8> PMID: 25972927; PubMed Central PMCID: PMC4429822.
36. Wong P, Hattangadi SM, Cheng AW, Frampton GM, Young RA, Lodish HF. Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. *Blood.* 2011; 118(16):e128–38. <https://doi.org/10.1182/blood-2011-03-341404> PMID: 21860024; PubMed Central PMCID: PMC3204918.
37. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature.* 2008; 454(7205):766–70. PMID: 18600261. <https://doi.org/10.1038/nature07107>
38. Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 2009; 37(Database issue):D32–6. Epub 2008/10/16. <https://doi.org/10.1093/nar/gkn721> PMID: 18927115; PubMed Central PMCID: PMC2686572.
39. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A.* 1998; 95(25):14863–8. PMID: 9843981; PubMed Central PMCID: PMC24541.
40. Jones BW, Pfeiffer RL, Ferrell WD, Watt CB, Marmor M, Marc RE. Retinal remodeling in human retinitis pigmentosa. *Exp Eye Res.* 2016; 150:149–65. Epub 2016/03/26. <https://doi.org/10.1016/j.exer.2016.03.018> PMID: 27020758; PubMed Central PMCID: PMC5031517.
41. Hackam AS, Strom R, Liu D, Qian J, Wang C, Otteson D, et al. Identification of gene expression changes associated with the progression of retinal degeneration in the rd1 mouse. *Invest Ophthalmol Vis Sci.* 2004; 45(9):2929–42. <https://doi.org/10.1167/iovs.03-1184> PMID: 15326104.
42. Watson AJ, Aragay AM, Slepak VZ, Simon MI. A novel form of the G protein beta subunit Gbeta5 is specifically expressed in the vertebrate retina. *J Biol Chem.* 1996; 271(45):28154–60. PMID: 8910430.
43. Nair KS, Balasubramanian N, Slepak VZ. Signal-dependent translocation of transducin, RGS9-1-Gbeta5L complex, and arrestin to detergent-resistant membrane rafts in photoreceptors. *Curr Biol.* 2002; 12(5):421–5. PMID: 11882295.
44. Watson AJ, Katz A, Simon MI. A fifth member of the mammalian G-protein beta-subunit family. Expression in brain and activation of the beta 2 isotype of phospholipase C. *J Biol Chem.* 1994; 269(35):22150–6. PMID: 8071339.

45. Zhang JH, Lai Z, Simonds WF. Differential expression of the G protein beta(5) gene: analysis of mouse brain, peripheral tissues, and cultured cell lines. *J Neurochem.* 2000; 75(1):393–403. PMID: [10854285](#).
46. André E, Conquet F, Steinmayr M, Stratton SC, Porciatti V, Becker-André M. Disruption of retinoid-related orphan receptor beta changes circadian behavior, causes retinal degeneration and leads to vacillans phenotype in mice. *EMBO J.* 1998; 17(14):3867–77. <https://doi.org/10.1093/emboj/17.14.3867> PMID: [9670004](#); PubMed Central PMCID: [PMCPMC1170722](#).
47. Liu H, Kim SY, Fu Y, Wu X, Ng L, Swaroop A, et al. An isoform of retinoid-related orphan receptor β directs differentiation of retinal amacrine and horizontal interneurons. *Nat Commun.* 2013; 4:1813. <https://doi.org/10.1038/ncomms2793> PMID: [23652001](#); PubMed Central PMCID: [PMCPMC3671912](#).
48. Fu Y, Liu H, Ng L, Kim JW, Hao H, Swaroop A, et al. Feedback induction of a photoreceptor-specific isoform of retinoid-related orphan nuclear receptor β by the rod transcription factor NRL. *J Biol Chem.* 2014; 289(47):32469–80. <https://doi.org/10.1074/jbc.M114.605774> PMID: [25296752](#); PubMed Central PMCID: [PMCPMC4239602](#).
49. Roof DJ, Hayes A, Adamian M, Chishti AH, Li T. Molecular characterization of abLIM, a novel actin-binding and double zinc finger protein. *J Cell Biol.* 1997; 138(3):575–88. PMID: [9245787](#); PubMed Central PMCID: [PMCPMC2141644](#).
50. Geisert EE, Lu L, Freeman-Anderson NE, Templeton JP, Nassr M, Wang X, et al. Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. *Mol Vis.* 2009; 15:1730–63. PMID: [19727342](#); PubMed Central PMCID: [PMCPMC2736153](#).
51. Stöhr H, Mohr N, Fröhlich S, Mehdi SQ, Bhattacharya SS, Weber BH. Cloning and characterization of WDR17, a novel WD repeat-containing gene on chromosome 4q34. *Biochim Biophys Acta.* 2002; 1579(1):18–25. PMID: [12401215](#).
52. Lu C, Ward PS, Kapoor GS, Rohle D, Turcan S, Abdel-Wahab O, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature.* 2012; 483(7390):474–8. Epub 2012/02/22. <https://doi.org/10.1038/nature10860> PMID: [22343901](#); PubMed Central PMCID: [PMC3478770](#).
53. Link S, Meissner M, Held B, Beck A, Weissgerber P, Freichel M, et al. Diversity and developmental expression of L-type calcium channel beta2 proteins and their influence on calcium current in murine heart. *J Biol Chem.* 2009; 284(44):30129–37. <https://doi.org/10.1074/jbc.M109.045583> PMID: [19723630](#); PubMed Central PMCID: [PMCPMC2781568](#).
54. Buraei Z, Yang J. The β subunit of voltage-gated Ca²⁺ channels. *Physiol Rev.* 2010; 90(4):1461–506. <https://doi.org/10.1152/physrev.00057.2009> PMID: [20959621](#); PubMed Central PMCID: [PMCPMC4353500](#).
55. Loeffen J, van den Heuvel L, Smeets R, Triepels R, Sengers R, Trijbels F, et al. cDNA sequence and chromosomal localization of the remaining three human nuclear encoded iron sulphur protein (IP) subunits of complex I: the human IP fraction is completed. *Biochem Biophys Res Commun.* 1998; 247(3):751–8. <https://doi.org/10.1006/bbrc.1998.8882> PMID: [9647766](#).
56. Shafit-Zagardo B, Kalcheva N. Making sense of the multiple MAP-2 transcripts and their role in the neuron. *Mol Neurobiol.* 1998; 16(2):149–62. <https://doi.org/10.1007/BF02740642> PMID: [9588626](#).
57. Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics.* 2013; 14:486. Epub 2013/07/18. <https://doi.org/10.1186/1471-2164-14-486> PMID: [23865674](#); PubMed Central PMCID: [PMCPMC3924432](#).
58. Sergouniotis PI, Davidson AE, Mackay DS, Lenassi E, Li Z, Robson AG, et al. Biallelic mutations in PLA2G5, encoding group V phospholipase A2, cause benign fleck retina. *Am J Hum Genet.* 2011; 89(6):782–91. Epub 2011/12/01. <https://doi.org/10.1016/j.ajhg.2011.11.004> PMID: [22137173](#); PubMed Central PMCID: [PMCPMC3234375](#).
59. Popova EY, Pinzon-Guzman C, Salzberg AC, Zhang SS, Barnstable CJ. LSD1-Mediated Demethylation of H3K4me2 Is Required for the Transition from Late Progenitor to Differentiated Mouse Rod Photoreceptor. *Mol Neurobiol.* 2016; 53(7):4563–81. <https://doi.org/10.1007/s12035-015-9395-8> PMID: [26298666](#).
60. Mo A, Luo C, Davis FP, Mukamel EA, Henry GL, Nery JR, et al. Epigenomic landscapes of retinal rods and cones. *Elife.* 2016; 5:e11613. Epub 2016/03/07. <https://doi.org/10.7554/eLife.11613> PMID: [26949250](#); PubMed Central PMCID: [PMCPMC4798964](#).
61. Gamsiz ED, Ouyang Q, Schmidt M, Nagpal S, Morrow EM. Genome-wide transcriptome analysis in murine neural retina using high-throughput RNA sequencing. *Genomics.* 2012; 99(1):44–51. Epub 2011/10/01. <https://doi.org/10.1016/j.ygeno.2011.09.003> PMID: [22032952](#); PubMed Central PMCID: [PMCPMC3392719](#).

62. Kim JW, Yang HJ, Brooks MJ, Zelinger L, Karakulah G, Gotoh N, et al. NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors. *Cell Rep.* 2016; 17(9):2460–73. <https://doi.org/10.1016/j.celrep.2016.10.074> PMID: [27880916](https://pubmed.ncbi.nlm.nih.gov/27880916/); PubMed Central PMCID: PMC5131731.
63. Ruzycski PA, Tran NM, Kefalov VJ, Kolesnikov AV, Chen S. Graded gene expression changes determine phenotype severity in mouse models of CRX-associated retinopathies. *Genome Biol.* 2015; 16:171. Epub 2015/09/01. <https://doi.org/10.1186/s13059-015-0732-z> PMID: [26324254](https://pubmed.ncbi.nlm.nih.gov/26324254/); PubMed Central PMCID: PMC4556057.