Identification and characterization of early human photoreceptor states and cell-state-specific retinoblastoma-related features

Short: Early photoreceptor states and cancer predisposition

Dominic W.H. Shayler,[1,2] Kevin Stachelek,[1,3] Linda Cambier,[1] Sunhye Lee,[1] Jinlun Bai, [1,2] Bhavana Bhat,[1] Mark W. Reid, [1] Daniel J. Weisenberger,[4,5] Jennifer G. Aparicio,[1] Yeha Kim,[1] Mitali Singh,[1] Maxwell Bay,[2] Matthew E. Thornton,[6] Eamon K. Doyle,[7,8] Zachary Fouladian,[1,2] Stephan G. Erberich,[7,8] Brendan H. Grubbs,[6] Michael A. Bonaguidi,[9] Cheryl Mae Craft,[10,11] Hardeep P. Singh,[1,11] David Cobrinik[1,4,5,11]*

[1] The Vision Center, Department of Surgery, and Saban Research Institute, Children's Hospital Los Angeles, Los Angeles, CA, USA.

[2] Development, Stem Cell, and Regenerative Medicine Program, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[3] Cancer Biology and Genomics Program, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[4] Department of Cancer Biology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[5] Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[6] Maternal-Fetal Medicine Division of the Department of Obstetrics and Gynecology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[7] Department of Radiology and The Saban Research Institute, Children's Hospital Los Angeles, Los Angeles, CA, USA.

[8] Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[9] Department of Development, Stem Cell, and Regenerative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[10] Department of Integrative Anatomical Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[11] USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

*Corresponding author: David Cobrinik, MD, PhD
Email: dcobrinik@chla.usc.edu

Abstract

Human cone photoreceptors differ from rods and serve as the retinoblastoma cell-of-origin, yet the developmental basis for their distinct behaviors is poorly understood. Here, we used deep full-length single-cell RNA-sequencing (scRNA-seq) to distinguish post-mitotic cone and rod developmental states and identify cone-specific features that contribute to retinoblastomagenesis. The analyses revealed nascent, immediately post-mitotic cone- and rod precursor populations characterized by higher THRB or NRL regulon activities, immature and maturing cone and rod precursors with concurrent cone- and rod-related gene and regulon expression, and distinct early and late cone and rod maturation states distinguished by maturation-associated declines in RAX regulon activity. Cell-state-specific gene expression features inferred based on full-length scRNA-seq were consistent with past single nucleus 3' RNA-seq analyses. Beyond the cell state characterizations, full-length scRNAseq revealed that both L/M cone and rod precursors co-expressed *NRL* and *THRB* RNAs, yet they differentially expressed functionally antagonistic *NRL* and *THRB* isoforms and prematurely terminated *THRB* transcripts. Moreover, early L/M cone precursors exhibited successive expression of several lncRNAs along with *MYCN*, which composed the seventh most L/M-cone-specific regulon, and *SYK*, which was implicated in the cone precursors' proliferative response to *RB1* loss. These findings reveal previously unresolved photoreceptor precursor states and suggest a role for early cone-precursor-intrinsic *SYK* expression in retinoblastoma initiation.

Impact Statement

Features acquired by early post-mitotic retinal cells underlie the distinct behaviors of rods and the cone cells of origin of retinoblastoma tumors.

Introduction

Vertebrate photoreceptors develop from optic vesicle retinal progenitor cells (RPCs) through progressive RPC lineage restriction, fate determination, and post-mitotic maturation[1,2]. While several transcription factors that govern these events have been identified, important aspects of photoreceptor fate determination and maturation remain unclear. For example, it is unclear if fate is determined in RPCs, where OTX2 and ONECUT1 are thought to control the post-mitotic expression of long- or medium-wavelength (L/M) cone determinant TRβ2 and rod determinant NRL[3], or is determined in post-mitotic photoreceptor precursors with concurrent TRβ2 and NRL expression[4]. Following fate commitment, post-mitotic developmental stages have been defined based on morphologic features and phototransduction-related gene or protein expression[5,6], but it is unclear if progression through such stages is subdivided into distinct cell states governed by unique transcription factor combinations or represents a developmental continuum.

An improved understanding of photoreceptor development may provide insight into the pathogenesis of retinal dystrophies, retinal degenerations, and the retinal cone precursor cancer, retinoblastoma[7,8]. In the latter case, L/M cone precursors lacking the retinoblastoma protein (pRB) were shown to proliferate in a manner dependent on the L/M-cone lineage factors RXRγ and TRβ2 and the intrinsically highly expressed MDM2 and MYCN oncoproteins, likely representing the first step of retinoblastoma tumorigenesis[8–11]. Similarly, retinoblastoma cell proliferation depends on RXRγ, TRβ2, MDM2, and MYCN[9], implying that intrinsic L/M-cone factors contribute to the oncogenic state. However, retinoblastoma cells also express rod lineage factor *NRL* RNAs, which – along with other evidence – suggested a heretofore unexplained connection between rod gene expression and retinoblastoma development[12,13]. Improved discrimination of early photoreceptor states is needed to determine if co-expression of rod- and cone-related genes is adopted during tumorigenesis or reflects the co-expression of such genes in the retinoblastoma cell of origin.

The cone precursors' propensity to form retinoblastoma is a human-specific feature whose study requires analysis of developing human retina[8]. Single-cell RNA-sequencing (scRNA-seq) is well suited to such analyses as it enables discrimination of cell type-specific fate determination and maturation features. scRNA-seq studies employing 3' end-counting have defined age-related post-mitotic transition populations, fate-determining features of post-mitotic photoreceptor precursors, and gene expression changes associated with the cone fate decision and early development[14–20]. However, 3' end-counting cannot be used to interrogate transcript isoforms, and the relatively low number of genes detected per cell limits the ability to distinguish closely related states in individual cells.

In this study, we sought to further define the transcriptomic underpinnings of human photoreceptor development and their relationship to retinoblastoma tumorigenesis. We generated deep, full-length single-cell transcriptomes of human retinal progenitor cells (RPCs) and developing photoreceptors from fetal week (FW) 13-19 retinae, with enrichment of rare cone precursor populations, and applied long-read cDNA sequencing, RNA velocity, pseudotemporal trajectory reconstruction, and single-cell regulatory network inference and clustering (SCENIC) to interrogate individual cell states. These analyses discriminated previously unresolved photoreceptor developmental states, identified photoreceptor precursor states with cone and rod-related RNA co-expression, uncovered cell type-specific expression of RNA isoforms of photoreceptor fate-determining genes, elucidated post-mitotic photoreceptor developmental trajectories, and revealed retinoblastoma cell-of-origin features that may contribute to retinoblastoma genesis.

Results

Regulon-defined RPC and photoreceptor precursor states

To interrogate transcriptomic changes during human photoreceptor development, dissociated RPCs and photoreceptor precursors were FACS-enriched from 18 retinae, ages FW13-19 (Figure S1A), and isolated using microfluidics or direct sorting into microliter droplets, followed by full-length cDNA synthesis, paired-end sequencing, and alignment to Ensembl transcript isoforms (Figure 1A). The FACS enrichment was based on a prior cone isolation method[10] but with wider gating to include rods and RPCs. After sequencing, we excluded all cells with <100,000 read counts and 18 cells expressing one or more markers of retinal ganglion, amacrine, and/or horizontal cells (*POU4F1*, *POU4F2*, *POU4F3*, *TFAP2A*, *TFAP2B*, *ISL1*) and concurrently lacking photoreceptor lineage marker *OTX2*. This yielded 794 single cells with averages of 3,750,417 uniquely aligned reads, 8,278 genes detected, and 20,343 Ensembl transcripts inferred (Figure S1A-C). Sequencing batches were normalized and transcriptomes clustered and visualized in uniform manifold approximation and projection (UMAP) plots that integrated cells across different retinae, ages, isolation methods, and sequencing runs (Figures 1B,C, S1D-F).

Low resolution Louvain clustering (level 0.4) generated six clusters that segregated into mostly distinct UMAP domains (Figure 1B). One cluster was comprised of RPCs and Müller glia (MG), with specific expression of *LHX2*, *VSX2*, *SOX2*, and *SLC1A3*, while five clusters were comprised of cells with photoreceptor features, with wide expression of *OTX2* and *CRX* and cluster-specific rod- and cone gene expression (Figures 1D, S2). In UMAP space, the cluster designated immature photoreceptor precursors (iPRPs) intermixed with the RPC/MG population, extended towards early L/M cone precursors, and was predominantly comprised of cells expressing the L/M cone determinant *THRB* (Figure 1B,D).

Two clusters highly expressing the rod determinant *NR2E3* were designated early-maturing rod (ER) and late-maturing rod (LR) (Figure 1B), based on the latter's increased

expression of rod phototransduction genes *GNAT1*, *CNGB1*, *PDE6G*, *GNGT1* and *RHO* (Figures 1D, S2C). Few rods were detected at FW13, whereas both early and late rods were detected from FW15-19 (Figure 1C), corroborating prior reports[15,20]. Differential expression analysis revealed other genes upregulated in late rods (*GNB1, SAMD7*, *NT5E*)[21,22] as well as the downregulated *CRABP2*, *DCT*, and *FABP7* (Figure S3A, Table S1). Genes upregulated in the LR cluster were enriched for photoreceptor and light sensing gene ontologies (Figure S3B) including spectrin binding, likely relating to proteins that control photoreceptor polarity and synapse formation[23,24], and purine biosynthesis and ribonucleotide metabolic processes, potentially related to the developing photoreceptors' high NAD+ requirements[25].

Cones segregated into distinct S- and L/M-cone clusters, with differential expression of cone subtype markers (*OPN1SW*, *THRB*), previously identified S-cone enriched genes (*CCDC136*, *UPB1*)[26–28]**,** novel S cone enriched genes (*MEGF10, NRXN3*, *ACKR3*), and the L/M cone transcription factor *ISL2*[15], among others (Figure 1B,D, S3C, Table S2). Gene ontology analysis did not reveal relevant terms enriched in S cones, whereas L/M cones were enriched for protein translation related ontologies (Figure S3D) due to increased expression of ribosomal protein genes *RPL23A*, *RPLP0*, *RPS19*, *RPS27*, *RPS27A*, *RPS29*, *RPS3A* (Table S2).

In UMAP space, the *THRB*+ L/M-cone cluster segregated into a large 366-cell proximal population and a 5-cell distal population inferred to represent early-maturing and late-maturing stages, respectively, based on the latter's increased expression of cone phototransduction genes *OPN1LW* (encoding L-opsin), *PDE6H*, and *GUCA1C* (Figure 1E, S3E,F), analogous to *RHO*, *PDE6G*, and *GNGT1* upregulation in late rods. L/M cone precursors from different age retinae occupied different UMAP regions, suggesting age-related differences in L/M cone precursor maturation (Figure 1C). Compared to the early L/M population, late L/M cones had upregulation of three M-opsin genes*, TTR*, encoding the retinol-binding protein transthyretin, *PCP4*, a small protein that binds calmodulin previously noted in foveal cones[29], and *MYL4*, a myosin light chain gene upregulated in retinal organoid L/M cones[28], among others (Figure

S3E,F, Table S3). The low proportions of *OPN1MW*+ and *OPN1LW*+ late-maturing L/M cones is consistent with a prior analysis of similar-age retinae and with the further upregulation of these proteins in later maturation[15].

We next asked whether similar distinctions between early-maturing and late-maturing L/M cone and rod precursors were observed in prior studies. Indeed, a 3' snRNA-seq analysis of ~ 220,000 retinal cells from post-conception week (pcw) 8 – 23 distinguished the *cone precursor* versus *ML cone* and the *rod precursor* versus *rod* subclasses (Zuo *et al*.[20]). (*N.b.*, we retain the *pcw* and *ML cone* terms of Zuo *et al.* when describing their data and the synonymous *FW* and *L/M cone* for our data to maintain continuity with past publications.) The Zuo *et al*. cone and rod precursor versus cone and rod photoreceptor comparisons were not strictly analogous to our early- maturing versus late-maturing precursor comparisons in that our early-maturing precursors excluded immature iPRP cells. Still, the comparisons revealed many of the same differentially expressed genes (Figures S3A versus S3G and S3E versus S3H; Tables S1, S3-S5).

To further interrogate cell identities, we used SCENIC to identify cluster-specific transcription factor regulons, which represent the overall expression of single transcription factors and their likely coregulated target genes[30]. The highest specificity regulons defined major cell populations including RPC/MG-specific E2F2, E2F3, VSX2, and PAX6; pan-photoreceptor NEUROD1, OTX2, and CRX; rod-specific NRL; and L/M cone-specific THRB and ISL2 (Figure 2, Table S6). SCENIC also distinguished ER and LR states via the latter's increased NRL, CRX, ATF4, and LHX3 and decreased HMX1 and RAX activities ($p < 0.0005$ for each, Dunn test). RAX activity also decreased in the 5-cell late-maturing L/M cone group (Figure 2B), supporting its distinct transcriptomic identity and suggesting a similar mode of late cone and late rod maturation. Additionally, iPRPs expressed their most specific regulons, LHX9 and OLIG2, at levels similar to RPC/MGs along with photoreceptor-related regulons, consistent with the transitional nature of this population (Figure 2A, Table S6). However, SCENIC did not

identify S cone-specific regulons, in keeping with the notion that S cones represent a default photoreceptor state induced by pan-photoreceptor factors such as OTX2, CRX, and NEUROD1 in the absence of NRL and THRB[31]. Thus, deep, full-length scRNA-seq enabled identification of regulons underlying RPC and developing photoreceptor states at the single cell level.

Differential expression of *NRL* and *THRB* isoforms in rod and cone precursors

Although cone and rod precursors segregated into distinct clusters, mRNAs encoding rod-determining factor NRL, L/M cone-determining factor TRβ, and cone marker RXRγ were co-expressed in both rod and cone precursor populations, with mean *NRL* expression only 4.3-fold higher in the ER vs LM cluster, mean *THRB* expression 5.0-fold higher in LM vs LR, and mean *RXRG* expression 4.3-fold higher in LM vs ER (Figure S2C,D and below). Cone *NRL* expression was unexpected given NRL's role in rod fate determination[28] and rod-specific NRL regulon activity (Figure 2A,D). Similarly, rod *RXRG* and *THRB* expression were unexpected given their roles in cone gene expression and fate determination[32,33] and L/M cone-specific THRB regulon activity (Figure 2A,D). Accordingly, we used full-length scRNA-seq data to determine if cone and rod precursors differentially express *NRL*, *THRB*, and *RXRG* transcript isoforms.

For *NRL*, three assigned transcript isoforms (*ENST00000397002*, *ENST00000561028*, and *ENST00000558280*) are predicted to encode the canonical full-length NRL protein (FL-NRL) (RefSeq NP_001341697.1), while two others (*ENST00000560550* and ENST00000396995) are previously uncharacterized transcripts predicted to use an alternative "P2" promoter and first exon, here termed exon 1T (Figure 3A-C). The novel transcripts are predicted to encode an N-terminally truncated NRL protein (Tr-NRL) retaining the leucine zipper DNA binding domain but lacking the minimal transactivation domain[34] (Figure 3C, Figure S4A,B). While all transcript isoforms were inferred to be more highly expressed in the ER rod cluster versus the LM cone cluster, the ratio of all FL-NRL:Tr-NRL transcripts was 2.9:1 in early rod precursors and 2.2:1 in late rod precursors, in contrast to 0.67:1 in L/M cone

precursors (Figure 3B). Consistent with the assigned isoform ratios, mean read coverage of the Tr-NRL-specific exon 1T was higher in S and LM cones while coverage of FL-NRL-specific exon 1 was higher in rods (Figure 3C, red *vs*. black arrowheads). Comparing the reads mapped to each first exon relative to total reads further confirmed that the Tr-NRL exon 1T predominated in individual cones whereas the FL-NRL exon 1 predominated in rods (Figure 3D). The cone cells' higher proportional expression of Tr-NRL first exon sequences was validated by RNA fluorescence *in situ* hybridization (FISH) of FW16 fetal retina in which NRL immunofluorescence was used to identify rod precursors, RXRγ immunofluorescence was used to identify cone precursors, and FISH probes specific to Tr-NRL exon 1T or to FL-NRL exons 1 and 2 were used to assess Tr-NRL and FL-NRL expression (Figure 3E,F).

While the Tr-NRL-encoding *NRL* isoforms were not to our knowledge previously described, another *NRL* isoform that initiated within exon 2 and lacked the NRL transactivation domain due to alternative exon 2 splicing, termed DD10 (Figure 3C), was previously identified in adult retina[36]. Concordantly, we detected reads spanning the unique DD10 splice junction, yet at lower levels than the unique Tr-NRL junction (5,942 versus 57,048).

As transcript isoforms inferred from short-read sequencing do not necessarily reflect the original transcript structures, we further defined the *NRL* isoforms expressed in early cone and rod precursors by performing 5' rapid amplification of cDNA ends (RACE) on the already generated single cell cDNA libraries from 23 early rod (ER) and 21 early L/M cone (LM) cells, followed by nanopore long-read sequencing of the pooled RACE products. The long-read sequencing revealed isoforms consistent with FL-NRL, Tr-NRL, DD10, and several other *NRL* isoforms with alternative transcription initiation and alternative splicing within exon 2 as well as within the Tr-NRL exon 1T (Figure 3G, Figure S4C). In keeping with the computationally assigned isoforms and differential exon usage, ER libraries had a higher proportion of FL-NRL exon 1 and exon 2 reads and LM libraries had a higher proportion of Tr-NRL exon 1T reads (Figure 3G, Figure S4C). Moreover, alternative splicing within NRL exon 2 was more prevalent

in LM libraries, affecting 6,788 of 29,177 (23 %) of exon 2 reads compared to 5,378 of 85,860 (6 %) in ER cells ($p<0.0001$; Chi-square with Yates correction), resulting in rarer full-length (exon 1-2-3) transcripts than inferred from short-read sequencing. Thus, long-read sequencing revealed cell type-biased *NRL* isoform expression with a preponderance of FL-NRL transcripts in early rods and disrupted FL-NRL and Tr-NRL transcript isoforms in L/M cones.

Despite our detection of L/M cone RNAs encoding Tr-NRL and FL-NRL, cone expression of NRL protein has not been reported. To assess endogenous Tr-NRL expression, we performed immunoblot analysis of CHLA-VC-RB-31 retinoblastoma cells[37], which were predicted to express FL-NRL and Tr-NRL transcripts in a cone-like 0.73:1 ratio, with an antibody that recognizes both FL-NRL and Tr-NRL proteins (Figure S5A,B). As in other retinoblastoma cells[13], FL-NRL increased in response to retinoic acid and proteasome inhibition, whereas Tr-NRL was not detected (Figure S5C). Similarly, Tr-NRL was not detected in EGFP+, RXRγ+ cones following lentiviral transduction of an explanted fetal retina with an EGFP-P2A-Tr-NRL cassette (Figure S5D-G). These findings suggest that Tr-NRL is poorly translated or unstable in most cone precursors.

As Tr-NRL might be expressed in contexts that were not examined in our analyses, we explored the function of the putative Tr-NRL protein. As both Tr-NRL and the previously described NRL DD10 lack the NRL transactivation domain and as DD10 interferes with FL-NRL transactivation[35], we examined if Tr-NRL similarly opposes FL-NRL transcriptional activity. Indeed, in luciferase reporter assays, Tr-NRL suppressed FL-NRL activation of a *PDE6B* promoter (Figure 3H).

For *THRB*, the most highly assigned transcript isoforms encoded the L/M cone-specific TRβ2 (*ENST00000280696)* and the more widely expressed TRβ1 (*ENST00000396671* and others) (Figure 4A-C, S6). While both TRβ1 and TRβ2 promote L/M cone fate determination[38], the isoform encoding TRβ2 predominated in L/M cones while those encoding TRβ1 predominated in early rods (Figure 4B). Moreover, late rods preferentially expressed *THRB*

exons 1-6 and the *THRB* 3' untranslated region (UTR), implying that RNAs encoding full-length TRβ proteins were rare (Figure 4B,C). Notably, a higher percentage of reads extended from the exon 4 and exon 6 splice donor sequences into the subsequent introns in LR versus LM cells (Figure 4D,E) ($p \leq 0.001$ for both, two-tailed Chi square test), suggesting that premature transcription termination (PTT) in introns 4 and 6 preferentially limits TRβ expression in the LR population. The inferred PTT events are consistent with structures of the assigned Ensembl isoforms (Figure 4B, S6).

To further evaluate PTT events, we used *THRB* 3' RACE and long-read sequencing on single cell cDNA libraries from 23 L/M cone cells and five LR cells selected for high *THRB* expression. RACE reactions were performed separately with primers complementary to the TRβ1-specific exon 4 and to the TRβ2-specific first exon (Figure 4F). Sequencing of LM and LR RACE products corroborated pronounced PTT in introns 4 and 6, with greater intron 6 PTT in LR versus LM cells (Figure 4F, Figure S7). However, we did not corroborate the late rods' proportionately higher intron 4 PTT, likely due to the small number of LR cells examined and heterogeneity in PTT frequency in individual cells. Long read sequencing also revealed PTT following the TRβ2-specific exon and a novel transcription-terminating exon following the canonical exon 5 observed solely in TRβ2 transcripts (Figure 4F, S7). 3' RACE transcripts rarely extended into the 3' UTR in LM or LR cells, suggesting that reads mapping to this region do not reflect protein-coding RNAs and confound the assessment of protein-coding *THRB* mRNA expression. These analyses demonstrate that *THRB* is regulated by multiple PTT events in rod and cone precursors as well as by cell type-specific promoter utilization and independent 3' UTR RNA expression.

For *RXRG*, short read sequencing reads were assigned to several isoforms that differed in their 5' promoter position and exon utilization (Figure S8A-C). However, 5' RACE and long-read sequencing did not support differential isoform expression (Figure S8D), and quantitative imaging revealed an average ~3.5-fold higher RXRγ protein in cones compared to rods (Figure

S8E). Thus, long-read sequencing clarified that *RXRG* expression is moderately higher in human cone versus rod precursors without evidence of cell-type specific isoforms.

Two post-mitotic immature photoreceptor precursor populations

To further define photoreceptor developmental states, we subdivided the initial six cell populations with higher resolution clustering (level 1.6), identified high resolution cluster-associated genes and regulons, and inferred each cell's rate and direction of transcriptomic change using RNA velocity[39] (Figure 5A,B). Increased clustering resolution divided the RPC/MG cluster into separate RPC and Müller glia (MG) groups, divided L/M cones into four subgroups (LM1 – LM4), which partially overlapped in UMAP space, and divided the iPRP cluster into two clusters here designated immature cone precursors (iCPs) and immature rod precursors (iRPs), the latter also drawing cells from the low-resolution ER cluster (Figure 5A,C). Similar clusters were observed at reduced k.param values that defines nearest neighbors.

The distinction between RPCs and MG was corroborated by the expression of known marker genes, with the RPC cluster having increased expression of cell cycle markers (*CCNE2*, *CCNA2*, *CCNB2*) and the cell cycle-associated *PBK* and *E2F7* (Figure 5D, S9A,B, Table S7). The MG cluster was spatially segregated, lacked expression of most cell cycle genes, and expressed genes known to be expressed in both populations (*CCND1*, *SLC1A3*, *PAX6*, *LHX2*, *VSX2*) as well as the MG marker *RLBP1*[40–42] (Figure 5C,D, S9A). Differential gene expression analyses revealed upregulation of cell cycle-related genes and G2/M checkpoint and E2F target ontologies in the RPC cluster but no significantly upregulated ontologies in the MG population (Figure S9B,C), consistent with early MGs resembling quiescent RPCs[43]. In contrast to other studies, we did not distinguish primary RPCs (PRPCs) from neurogenic RPCs (NRPCs)[14,15,20], likely due to the underrepresentation of RPCs in our dataset.

The high resolution iCP and iRP clusters included 'RPC-localized' cells positioned adjacent to RPCs and 'non-RPC-localized' cells adjacent to the LM and ER clusters in UMAP

space (Figure 5C). RNA velocity suggested that the RPC-localized cell groups flowed towards the larger early maturing L/M cones and rods, respectively (Figure 5B,C). Compared to RPCs and MG, RPC-localized iCP and iRP cells had minimal expression of cyclin RNAs or RPC markers *PAX6*, *LHX2*, and *VSX2*, and had increased expression of photoreceptor determinants *OTX2* and *CRX* (Figure 5D), consistent with their being immediately post-mitotic photoreceptor precursors. iCP cells in this region upregulated the early cone cell fate determinant *ONECUT1*[3,44], the L/M cone determinant *THRB*, and the cone differentiation marker *GNAT2*, whereas iRP cells trended towards higher expression of the rod determinant *NR2E3* (Figures 5D, S10A); however, such RNAs may be more highly expressed in cone- and rod-fated NRPCs than in the unspecified RPCs in our study. Notably, *THRB* and *GNAT2* expression did not significantly change while *ONECUT1* declined in the subsequent non-RPC-localized iCP and LM1 stages, whereas *NR2E3* and *NRL* dramatically increased on transitioning to the ER state (Figure S10A).

In the RPC-localized region, iCPs had higher *ONECUT1*, *THRB*, and *GNAT2*, whereas iRPs trended towards higher *NRL* and *NR2E3* ($p= 0.19$, $p=0.054$, respectively; Figure 5D). While we detected differential expression when selectively examining these genes of interest, no genes were differentially expressed (Padj < 0.05) in a transcriptome-wide comparison, likely due to the lack of statistical power given the small number of cells examined.

To our knowledge, past studies have not distinguished immature cone and rod precursors (*i.e.*, iCPs and iRPs) from the subsequent maturing cone and rod precursor states. To explore whether early cone and rod precursors with similar properties are present in droplet-based scRNA-seq studies, we examined gene expression in spatiotemporally segregated cone and rod precursor populations as well as in cone and rod NRPCs in the Zuo *et al*. 3' snRNA-seq dataset[20]. This revealed that cone NRPCs and the earliest (pcw 8-10) cone precursors had high *ONECUT1* and *THRB* yet minimal *NR2E3* (Figure S10B). Similarly, rod NRPCs had high *NRL* and *NR2E3*, yet it was not possible to examine the earliest (pcw 10-13) rod precursors, as many

had cone-like gene expression profiles, suggesting they may have been mis-assigned (Figure S10B, *grey boxes*). Further evaluation suggested misassignment of a smaller proportion of rod and cone precursors at other ages (Figure S11). In particular, cells potentially mis-assigned as rod precursors were located in the cone precursor UMAP region and highly expressed *THRB* (Figure S11D, **), while cells potentially mis-assigned as cone precursors were located in the rod precursor UMAP region and highly expressed *NR2E3* (Figure S11E, *). When considering only the main cone and rod precursor UMAP regions, early (pcw 8-13) cone precursors expressed *THRB* and lacked *NR2E3* (Figure S11D,E, blue arrows), while early (pcw 10-15) rod precursors expressed *NR2E3* and lacked *THRB* (Figure S11D,E, red arrows), similar to RPC-localized iCPs and iRPs in our study (Figure 5D). We conclude that Louvain clustering of deep full-length scRNA-seq distinguished immediately post-mitotic iCP and iRP populations with features similar to early cone and rod precursors in a large 3' snRNA-seq dataset.

In a further comparison, we noted lower detection of *GNAT2* in L/M cone precursors and lower detection *CCNE2*, *CCNA2*, and *CCNB2* in RPCs in 3' snRNA-seq versus full-length scRNA-seq datasets (Figure S10A,B). This suggests that deep full-length scRNA-seq enabled more sensitive detection of certain cell-type-specific RNAs along with improved discrimination of immature cone versus rod precursor states.

SCENIC regulons further clarified the identities of the four RPC-localized clusters. The MG cluster was best specified (*i.e.*, had highest regulon specificity scores) by PAX6 and SOX9, both previously described in RPCs and MGs[40,45,46], whereas RPCs were best specified by E2F2 and E2F3 (Figure 5E). These RPC and MG regulons were low or absent in the early post-mitotic iCPs and iRPs, suggesting they undergo an abrupt cell state change. Moreover, RPC-localized iCPs had greater THRB regulon signal than RPC-localized iRPs, while RPC-localized iRPs had higher NRL regulon signal ($p<0.05$ for both) (Figure 5E). Interestingly, RPC-localized and/or non-localized iCPs also had higher activity of the pan-photoreceptor regulons LHX3, OTX2, and NEUROD1 compared to their iRP counterparts (Figure S10C,D). Finally, activities of the cone-

specific THRB and ISL2 regulons, the rod-specific NRL regulon, and the pan-photoreceptor LHX3, OTX2, CRX, and NEUROD1 regulons increased to varying extents on transitioning from immature iCP or iRP states to the early-maturing LM1 or ER states (Figure 10C).

Early cone and rod precursors with rod- and cone-related RNA co-expression

In addition to the immediately post-mitotic RPC-localized iCPs, the iCP cluster included cells bridging the UMAP region between early maturing cones and early maturing rods (Figure 6A). Many iCP and ER cells in this bridge region expressed RNAs encoding cone markers (*GNAT2*, *THRB*), rod markers (*GNAT1*, *NR2E3*), or, in many cases, both (Figure 6B), suggestive of a proposed hybrid cone/rod precursor state more extensive than implied by the co-expression of different *THRB* and *NRL* isoforms[4].

To determine whether the intermixed cone and rod gene expression reflects a hybrid cone-rod precursor state, we examined bridge region NRL and THRB regulon activities, which embody the overall transcriptomic effects of these factors. Indeed, in the bridge region, cells with NRL and THRB regulon signals intermixed, and some cells showed signals for both (Figure 6C, arrows). However, iCP cells had low NRL regulon z-scores similar to L/M cones and most had THRB z-scores above that of adjacent ER cells (Figure 6D). The difference between regulon signals increased in rod and cone clusters outside the bridge region as NRL or THRB regulons became more dominant (Figure 6D).

To identify cells with early cone- and rod-related RNA co-expression in the developing retina, we performed multiplex RNA FISH for early cone marker *GNAT2* and rod marker *NR2E3* combined with immunofluorescence (IF) staining for RXRγ, which has high expression in outermost neuroblastic layer (NBL) cone precursors and low expression in middle NBL rod precursors (as earlier shown in Figure S8E), and for NR2E3, which is detected solely in rods. To infer the spatiotemporal pattern of such expression across human retinal development, we examined *GNAT2* and *NR2E3* RNA co-expression in RXRγ+ cone precursors in the outermost

NBL and in RXRγ+ rod precursors in the middle NBL across 13 regions of a FW14 retina section (Figures 7; see Figure S12 for IF + FISH images). Limiting our analysis to the outer and middle NBL allowed us to disregard RXRγ+ retinal ganglion cells in the retinal ganglion cell layer or inner NBL.

The analyses revealed that most of the far peripheral (hence, nascent) outer NBL RXRγ+ cone precursors (Figure 7A,B, regions 2 and 13) lacked detectable *GNAT2* and *NR2E3* RNA (green bars in Figure 7C), whereas those in the more mature central retina (regions 3-12) were uniformly *GNAT2*+ (yellow and red bars in Figure 7C). However, starting with regions 3 and 11, some outermost NBL *GNAT2*+ cones were also *NR2E3*+ (red bars in Figure 7C). The proportion of outermost NBL cells that co-expressed *GNAT2* and *NR2E3* RNA increased in the more central retina (regions 5-10), yet all of these cells had strong RXRγ staining and lacked NR2E3 protein, consistent with their having a cone identity (illustrated in Figure 7E). In contrast, most middle NBL RXRγ+ cells had either a low number of *NR2E3* RNA puncta without NR2E3 protein (*NR2E3*$^{+}$ / NR2E3^{-}, magenta bars in Figure 7D), likely in nascent rods, or prominent *NR2E3* RNA with NR2E3 protein (*NR2E3*$^{+}$ / NR2E3^{+}, purple bars in Figure 7D), in maturing rods (Figure 7D, E). However, we did not detect *GNAT2* puncta in middle NBL RXRγ+, *NR2E3*+ cells, including the most peripheral RXRγ+ cells with low-level *NR2E3* RNA (Figures 7E, S12). In summary, most photoreceptor precursors with *GNAT2* and *NR2E3* RNA co-expression had high RXRγ, no detectable NR2E3 protein, and outermost NBL positions expected of cone precursors. This supports the notion that, in our scRNA-seq analyses, bridge region iCP cells with combined *GNAT2* and *NR2E3* RNA expression are likely cone-directed.

Cone and rod precursor populations that co-express cone and rod marker genes were also evident in the Zuo *et al*. 3' snRNA-seq dataset. While the earliest cone precursors were *NR2E3*-negative and the earliest rod precursors were *THRB*-negative (as described above), starting at pcw 15, later cone precursors expressed *NR2E3* (Figure S11E, green arrows) and

later rod precursors expressed *THRB* (Figure S11D, purple arrows). *THRB* was highest in the most distal rod precursors (yet still lower than in cone precursors), corroborating our findings in Figure 1D and Figure 4B-C. In contrast, in the 3' snRNA-seq analysis, *GNAT2* expression was sparse in rod precursors of all ages (Figure S11F), consistent with the lack of *GNAT2* in rod precursors in our RNA FISH analyses and implying that different cone- and rod-related genes have different tendencies to be expressed in the other photoreceptor precursor type.

Thus, a 3' snRNA-seq analysis confirmed the initial production of immature photoreceptor precursors with either L/M cone-precursor-specific *THRB* or rod-precursor-specific *NR2E3* expression, followed by lower-level co-expression of their counterparts, *NR2E3* in cone precursors and *THRB* in rod precursors. However, in the Zuo *et al*. analyses, the co-expression was first observed in well-separated UMAP regions, as opposed to a region that bridges the early cone and early rod populations in our UMAP plots. These findings are consistent with the notion that cone- and rod-related RNA co-expression begins in already fate-determined cone and rod precursors, and that such precursors aberrantly intermixed in our UMAP 'bridge region' due to their insufficient representation in our dataset.

Developmental expression of photoreceptor precursor markers and fate determinants

To further assess relationships between iCP subpopulations, we examined the expression and UMAP distributions of the four iCP marker genes identified in Figure S9A. Among these genes, *CHRNA1* was mainly expressed in bridge region iCP and ER cells, whereas *ONECUT1* was biased to cone-directed iCPs, lncRNA *CTC-378H22.2* was expressed in a narrow zone of *THRB+, GNAT2+* iCPs, and *S100A6* was more widely expressed (Figure 8A). iCP cells also expressed *ATOH7* and *DLL3*, which were previously proposed to define transitional photoreceptor precursors and promote cone fate determination[14–16]. *ATOH7* was largely restricted to RPC-localized and bridge region iCPs whereas *DLL3* was broadly expressed similar to *S100A6* (Figure 8A).

A similar pattern was seen in the Zuo *et al*. 3' snRNA-seq dataset[20], where *CHRNA1*, *ONECUT1*, *S100A*, *ATOH7*, and *DLL3* were most highly expressed in the youngest and developmentally earliest ML cone and rod precursors and persisted to various extents during cone and rod maturation, whereas *CTC-378H22.2* (ENSG00000259436) had more restricted and cone-specific expression (Figure S13A). However, the early cone and rod precursor UMAP regions did not adjoin one another (in contrast to our UMAP bridge region), suggesting that the cone and rod precursors expressing the above markers had begun their distinct trajectories and that their juxtaposition in our UMAP analyses is spurious. Also, different iCP markers had different spatiotemporal expression: *CHRNA1* and *ATOH7* were most prominent in peripheral retina with *ATOH7* strongest at pcw 10 and *CHRNA1* strongest at pcw 13; *CTC-378H22.2* was prominently expressed from pcw 10-13 in both the macula and the periphery; and *DLL3* and *ONECUT1* showed the earliest, strongest, and broadest expression (Figure S13B). The distinct patterns suggest spatiotemporally distinct roles for these factors in cone precursor differentiation.

As *CHRNA1* appeared to be the most specific marker of both early cone and early rod precursors, we evaluated whether *CHRNA1* RNA marked specific cone and rod populations in the developing retina using RNA-FISH combined with RXRγ and NRL immunofluorescent staining. In a FW12 retina, we detected highest *CHRNA1* in the earliest, most peripheral NRL^+ rods and $RXR\gamma^+$ cones and fewer *CHRNA1*+ cells in more mature, central regions ($p < 0.005$ in cones) (Figure 8B-D), consistent with its expression in early cone and rod precursors.

To identify factors that regulate transcriptomic states during cone cell fate determination, we examined the UMAP distributions of the most iCP-specific transcription factor regulons, OLIG2 and LHX9 (Figure S10D, Table S8). As Olig2 was previously detected in mouse RPCs in which Onecut1 enabled *Thrb* expression and cone fate determination[3,47], the OLIG2 regulon was expected to be most active in RPCs and in early, immediately post-mitotic iCPs. While the

OLIG2 and LHX9 regulons were indeed active in RPCs, both were also active in a narrow zone of *ONECUT1*+ iCPs positioned farthest from S cones and immediately preceding the upregulation of the THRB regulon, a location with sparse *OLIG2* RNA expression (Figure S14). These data indicate that the OLIG2 and LHX9 regulons are active in and potentially relevant to human L/M cone fate determination in a post-mitotic iCP subpopulation.

An early L/M cone trajectory marked by successive lncRNA expression

After fate commitment, L/M cones initiate a maturation process with upregulation of RNAs and proteins related to phototransduction, axonogenesis, synaptogenesis, and outer segment morphogenesis[5,6]. To evaluate whether early L/M cone maturation is comprised of distinct transcriptomic cell states, we assessed marker gene and regulon differences between high resolution clusters LM1, LM2, LM3, and LM4. While clusters LM1-4 were distinguished by sequential increasing expression of *ACOT7*, *RTN1*, *PDE6H*, *OLAH*, and *NPFF*, they showed only subtle differences among the three regulons with highest LM1-4 specificity scores, THRB, ISL2, and LHX3 (Figure 9A,B). The lack of cluster-specific marker genes and regulons suggest that LM1-LM4 represent different stages of a graded maturation process.

We further evaluated gene expression changes in maturing cones by pseudotemporally ordering iCP and L/M cone precursors (Figure 9C). This identified 967 pseudotime-correlated genes (q-value < 0.05, expression > 0.05 in >5% of cells; Table S9) in seven gene modules (Figure S15). Among the top 20 pseudotime-correlated genes in each module, we identified four lncRNAs that were sequentially expressed (Figure 9D). To determine if these lncRNAs distinguish developmentally distinct maturing cones *in vivo*, we probed their expression via multiplex RNA FISH in co-stained RXRγ+ cone precursors across a FW16 retina (Figure 9E,F). Quantitation of FISH puncta defined four cone maturation zones based on expression peaks and significant count differences for each lncRNA as color-coded in Figure 9G: the most peripheral cones with high *CTC-378H22.2* and *HOTAIRM1* (blue), peripheral cones with high

HOTAIRM1 only (green), cones with low expression of all four lncRNAs (red), and parafoveal/foveal cones with upregulated *RP13143G15.4*, C*TD-2034I21.2*, and *CTC-378H22.2* (purple) (Figure 9G). The detection of foveal *CTC-378H22* by ISH but not in late LM transcriptomes may relate to a lack of the most mature cones in our scRNA-seq analyses. These data support the concept that maturing L/M cones sequentially express specific lncRNAs as they develop.

Cone-intrinsic SYK expression associated with the proliferative response to pRB loss

We next assessed whether early-maturing cone transcriptomic features are conducive to the proliferative response to pRB loss[10,11]. Given the high transcriptomic similarity across the L/M cone population, we compared gene expression between all early-maturing L/M cones and early-maturing rods. This identified 422 genes upregulated and 119 downregulated in cones ($p<0.05$, $\log2FC>|0.4|$) (Figure S16A, Table S10). Among cone-enriched genes, the top three enriched ontologies related to translation initiation, protein localization to membrane, and MYC targets (Figure S16B). The upregulation of MYC target genes was of interest given that many MYC target genes are also targets of MYCN, that MYCN protein is highly expressed in maturing (ARR3+) cone precursors but not in NRL+ rods (Figure 10A), and that MYCN is critical to the cone precursor proliferative response to pRB loss[8–10]. Indeed, whereas *MYC* RNA was not detected, the LM cone cluster had increased *MYCN* RNA ($\log_2FC = 0.54$) and MYCN regulon activity, representing the seventh highest LM cluster regulon specificity score (Figure 10B,C, Table S6).

Among other differentially expressed genes, we noted the L/M cone-specific upregulation of *SYK* (Figure 10D, Table S10), which encodes a non-receptor tyrosine kinase. Whereas SYK was previously implicated in retinoblastoma genesis and proposed to be induced in response to pRB loss[48], its expression was not previously reported in developing fetal retina. Indeed, our scRNA-seq analyses revealed that *SYK* RNA expression increased from the iCP

stage through cluster LM4, in contrast to its minimal expression in rods (Figure 10E). Moreover, *SYK* expression was abolished in the five-cell group with properties of late-maturing cones (characterized in Figure 1E), here displayed separately from the other LM4 cells and designated LM5 (Figure 10E). Similarly, immunohistochemical staining revealed high SYK protein expression in immature (ARR3-) and early-maturing (ARR3+) cones from the retinal periphery to the maturing foveal cones at FW16, while SYK was not detected in the most mature foveal cones at FW18 (Figure 10F). The loss of SYK protein and RNA expression with cone maturation is consistent with the lack of SYK in cones of normal retina adjacent to a retinoblastoma tumors[48] and implies that SYK is a defining feature of the human early cone precursor state. While *MYCN* RNA was also preferentially expressed in early cones, it did not increase as much relative to RPCs nor decline as much in late-maturing cone precursors when compared to *SYK* RNA dynamics (Figure 10E).

To determine if SYK might contribute to retinoblastoma initiation, dissociated fetal retinal cells were transduced with an *RB1*-directed shRNA (shRB1-733) known to induce cone precursor proliferation[10], treated with the selective SYK inhibitor GS-9876[49] for 12 days, and examined for cone precursor cell cycle entry by co-staining for RXRγ and Ki67. GS-9876 treatment suppressed the proliferative response to pRB knockdown at all concentrations from 1.0 – 5.0 μM (Figure 10G), consistent with the notion that cone precursor intrinsic SYK activity contributes to the proliferative response to pRB loss and is retained in retinoblastoma cells (Figure 10H). However, given potential SYK inhibitor off-target effects, validation of the role of SYK in retinoblastoma initiation will require gene ablation studies.

Discussion

This study evaluated cell state changes associated with human cone and rod photoreceptor development using deep full-length scRNA-seq. Whereas prior scRNA-seq studies provided insights into RPC fate determination and trajectories using 3' end counting[14–18,20,26], deep full-

length sequencing enabled more precise resolution of cell states and identification of cell type-specific regulon activities and transcript isoforms. Additionally, our cell enrichment strategy provided insight into the transcriptomic profiles and potential cancer-predisposing features of developing cones, a rare population with unique cancer cell-of-origin properties[8,10,11].

One advantage of full-length scRNA-seq is that it enables the detection of differential transcript isoform expression. Here, we show that the rod determinant *NRL*, the L/M-cone determinant *THRB*, and the cone marker *RXRG* are all co-expressed in cone and rod precursors at the RNA level yet use distinct cell type-specific mechanisms to enable appropriate differential protein expression. Long-read cDNA sequencing revealed that L/M cones preferentially express novel *NRL* transcript isoforms predicted to encode a truncated NRL protein (Tr-NRL) lacking a transactivation domain as well as *NRL* isoforms with intra-exon-2 transcription initiation and alternative splicing. As with DD10[35], Tr-NRL opposed transactivation by full-length NRL and thus may suppress rod-related transcription. However, our inability to detect intrinsic Tr-NRL or to overexpress Tr-NRL in cone precursors suggests there are additional layers of regulation, possible effects of untranslated Tr-NRL RNA, and alternative contexts in which these isoforms act. A similar assortment of *NRL* isoforms were expressed in rods, yet a far higher proportion of rod *NRL* transcripts encoded full-length (FL-) NRL protein. Thus, the L/M cone precursors' greater use of the Tr-NRL first exon and *NRL* exon 2 disruptions reveal a multipronged approach to suppress FL-NRL function while downregulating overall *NRL* RNA by only ~ 4-fold.

Similarly, we uncovered cell type-specific differences in *THRB* transcript isoforms, albeit generated through premature transcription termination (PTT) and 3' UTR expression in late rod precursors. PTT is widely used to govern expression of transcription regulators[50], and while PTT-shortened *THRB* transcripts had been identified[51], their retinal cell specificity was not previously recognized. In contrast, the expression of *THRB* 3' UTR sequences independent of

THRB protein-coding sequences was not previously reported and may enable additional regulation[52]. As a general matter, the expression of 3' UTR transcripts that are detached from protein-coding sequences may be inferred from full-length scRNA-seq but may be misinterpreted to represent protein-coding transcripts and thus confound the interpretation of scRNA-seq performed with 3' end-counting. To address this issue, our publicly available Shiny app displays exon coverage for all genes expressed in the RPC and photoreceptor clusters in this study.

One limitation of these studies is that individual cells may express a small and varying spectrum of transcript isoforms. While this variability was mitigated by combining cDNAs of cells deemed to be in similar states based on short-read sequencing, analysis of more cells might better reveal the spectrum of isoforms expressed by specific cell populations.

Understanding human photoreceptor development requires the identification of photoreceptor lineage developmental states and elucidation of mechanisms underlying their transitions. While cell states may be defined in various ways, at the transcriptome level they are perhaps best defined by unique combinations of transcription factor regulon activities[30]. Our deep sequencing approach uncovered discrete cell states distinguished by gene expression as well as by regulon activities. For example, early- and late-maturing rod populations were distinguished by the latter's increased expression of phototransduction genes together with increased activity of the NRL, ATF4, CRX, and LHX3 regulons and decreased HMX1 and RAX regulons (Figure 2). Similarly, L/M cones formed an early-maturing cone cluster characterized by high THRB and ISL2 regulons, consistent with THRB and ISL2 driving L/M cone identity[4,15,53], and a late-maturing cone group with decreased RAX activity. Downregulation of the RAX regulon in late-maturing rod and cone precursors is consistent with decreasing RAX protein during photoreceptor maturation and with RAX-mediated suppression of the cone opsin and rhodopsin promoters[54,55].

Increased clustering resolution further divided the L/M cone cluster into four subgroups with subtle maturation-associated changes in marker gene expression and regulon activity, consistent with LM1-4 comprising different stages of a graded maturation process (Figure 9). However, trajectory analyses revealed successive expression of lncRNAs that was validated in developing retinal tissue. Three of these lncRNAs (*HOTAIRM1*, *CTD-2034I21.1* (neighbor gene to *CTD-2034I21.2*), and the mouse ortholog of *CTC-378H22.2* (D930028M14Rik)) were previously observed in cone scRNA-seq[15,17,56]. Their sequential expression in the absence of discrete regulon changes suggests that they demarcate early L/M cone precursor substates.

High resolution clustering also segregated immature cone precursor (iCP) and immature rod precursor (iRP) populations, which were further partitioned according to their UMAP positions. *RPC-localized* iCP and iRP cells (those positioned adjacent to RPCs and MG in UMAP plots) lacked cell cycle-related gene expression, showed higher L/M cone-like (THRB) or rod-like (NRL) gene expression and regulon activities, respectively, and had RNA velocities directed toward distinct cone and rod populations (Figure 5). iCPs also upregulated core photoreceptor regulons OTX2 and NEUROD1 to a greater extent than iRPs (Figure S10C), supporting their distinct trajectories. Prior droplet-based scRNA-seq analyses have not to our knowledge distinguished immediately post-mitotic immature cone and rod precursors from the subsequent early-maturing cone and rod precursor stages.

Whereas deep full-length scRNA-seq enabled discrimination of immature cone and rod precursors, 3' scRNA-seq and snRNAs-seq enabled discrimination of the PRPC and cone- and rod-fated NRPCs [14,15,19 20] (Figure S17A). Combining these observations supports a model in which cone- and rod-fated NRPCs give rise to immediately post-mitotic (*i.e.*, *nascent*) and subsequent immature cone and rod precursors (iCPs and iRPs), which transition to early-maturing and then late-maturing L/M cone and rod precursors, with each state having unique gene expression and regulon properties (Figure S17B,C). It will be important to corroborate this model with lineage tracing, to extend the model to S cones, and to determine whether late-

maturing cone and rod states in fetal retina are distinct from mature post-natal cone and rod photoreceptors at the transcription factor regulon level.

Although not identified as a distinct cluster, our analyses also revealed cone and rod precursor subpopulations that co-express rod and cone genes and regulons (Figure 6, 7, S11, S12, S17C). In both the current full-length scRNA-seq and a recent 3' snRNA-seq analysis[20], the earliest cone and rod precursors lacked such co-expression, implying that this property is acquired after cone versus rod fate-determination. In support of this notion, *NR2E3* RNA was detected in more mature central retina cone precursors but not in nascent peripheral cone precursors in fetal retina tissue (Figure 7). However, more information is needed to assess whether such co-expression serves a developmental purpose. For example, it is unknown if cone precursor *NR2E3* RNA ever produces NR2E3 protein, as in zebrafish retina, which could suppress cone-related gene expression[57] and delay terminal differentiation. Similarly, it is unclear if the *THRB* RNA expressed in rod precursors – which is largely truncated or comprised of non-coding 3' UTR sequences (Figure 4) – has direct RNA effects[52].

Our characterization of cone and rod-related RNA co-expression may help resolve questions about the retinoblastoma cell of origin. Past studies suggested that retinoblastoma cells co-express RNAs associated with rods, cones, or other retinal cells due to a loss of lineage fidelity[12]. However, the early L/M cone precursors' expression of *NR2E3* and *NRL* RNAs suggest that their presence in retinoblastomas[12,13] reflects their normal expression in the L/M cone precursor cells of origin. This idea is further supported by the retinoblastoma cells' preferential expression of cone-enriched *NRL* transcript isoforms (Figure S5B).

Our analyses also refine understanding of gene expression in the earliest cone and rod precursors. Early cone and rod precursors shared expression of neurogenic precursor markers *ATOH7*, *DLL3*, and the new marker *CHRNA1*, albeit with higher expression of each in the cone lineage and preferential expression of *ATOH7* and *CHRNA1* in the retinal periphery (Figures 8, S13). *ATOH7*, *DLL3*, and *CHRNA1* RNAs persist in some maturing cone and rod precursors,

whereas *CTC-378H22.2* more precisely marks iCPs. iCPs with velocity directed towards early-maturing L/M cones expressed *THRB* and *ONECUT1* RNAs and had OLIG2 regulon activity (Figure S14); as these elements were proposed to promote cone fate in lineage-restricted RPCs[3], their expression in iCPs suggests a similar role, such as in L/M cone fate determination, in post-mitotic cells.

We also mined gene expression differences in early cone versus rod photoreceptors to identify factors that impact human cone sensitivity to *RB1* loss[10,11]. We detected upregulation of genes targeted by MYC, some of which are similar to those targeted by MYCN[58], along with upregulated *MYCN* RNA, MYCN protein, and MYCN regulon activity. As MYCN is required for the proliferative response to pRB loss[10] and triggers retinoblastoma when amplified and overexpressed[8], these findings suggest that increased *MYCN* RNA expression and regulon activity contribute to pRB-deficient cone precursor proliferation and retinoblastoma genesis. We further found that cone precursors highly express SYK, an oncoprotein previously detected in human retinoblastomas and *RB1*-null retinal organoids but not in healthy tumor-associated retina[48,59]. The high SYK expression preceding early cone precursor maturation and its loss during late maturation implies that high-level SYK expression is a retinoblastoma cell-of-origin-specific feature. Moreover, the pRB-depleted cone precursors' sensitivity to a SYK inhibitor suggests that native SYK expression rather than *de novo* induction contributes to the cone precursors' initial proliferation (Figure 10H), although genetic ablation of *SYK* is needed to confirm this notion.

In summary, through deep, full-length RNA sequencing, we identified photoreceptor cell type-specific differences in transcript isoform expression and post-mitotic photoreceptor precursor states with distinctive gene expression and regulon activities. The discrimination of these states enabled the identification of developmental stage-specific cone precursor features associated with the cone precursors' predisposition to form retinoblastoma tumors.

Materials and Methods

Primary human tissue

Human fetal samples were provided by Advanced Bioscience Resources, Inc., Alameda, CA, or collected under Institutional Review Board approval at the University of Southern California (protocol HS-13-00399), and Children's Hospital Los Angeles (protocol CHLA-14-00122). Following the patient decision for pregnancy termination, patients were offered the option of donation of the products of conception for research purposes, and those that agreed signed an informed consent. This did not alter the choice of termination procedure, and the products of conception from those who declined participation were disposed of in a standard fashion. Fetal age was determined according to the American College of Obstetrics and Gynecology guidelines[60]. Each retina subjected to sequencing was from a different donor.

Retinal dissociation and RNA-sequencing

Single cell isolation

Retinae were dissected while submerged in ice-cold phosphate-buffered saline (PBS) and dissociated as described[10]. Briefly, retina was removed and placed in 200 µl Earle's balanced salt solution (EBSS) in a 6-well plate on ice. 2 ml of 10 u/ml papain (Worthington LK003176) solution was added and then incubated at 37°C for 10 min, followed by pipetting with a 1000 µl pipet tip to break up large pieces and additional 5 min incubations until cells were dissociated to single cell level. An additional 1-2 ml papain was added after 20 min if dissociation was insufficient (~40 min total). Retina culture media [Iscove's Modified Dulbecco's Medium with glutamine (Corning, #12440046), 10% fetal bovine serum (FBS, FB-01, Omega Scientific), 0.28 U/mL insulin (Eli Lilly NDC 0002-0213-01), 55 µM β-mercaptoethanol (Fisher Scientific, # 21985023), penicillin and streptomycin] (Fisher Scientific, # MT30002CI) was used to stop enzyme activity and cells were centrifuged in 14 ml round bottom tubes at 300 x g, 4°C for 10 min. Supernatants were centrifuged at 1100 x g for an additional 3 min. After resuspension in

Hank's balanced salt solution (Fisher Scientific, 14025092), a 10 μl volume of cells was combined with 10 μl trypan blue and counted using a hemocytometer.

FACS isolation of single cells and cDNA synthesis

Single cells were FACS-isolated as described[10] with differences as follows. Dissociated cells were centrifuged and resuspended in a 5% fetal bovine serum in PBS (FBS/PBS) with CD133-PE (Miltenyi Biotec 130-113-108), CD44-FITC (BD Pharmingen 555478) and CD49b-FITC antibodies (BD Pharmingen 555498) to a final concentration of 10,000 cells/μl. After incubation at room temperature (RT) for 1 h, samples were diluted in 5% FBS/PBS, centrifuged as above, and resuspended in 300-400 μl 1% FBS/PBS containing 10 μg/ml 4′,6-diamidino-2-phenylindole (DAPI). Single cells were sorted on a BD FACSAria I at 4°C using 100 μm nozzle in single-cell mode into each of eight 1.2 μl lysis buffer droplets on parafilm-covered glass slides, with droplets positioned over pre-defined marks (S. Lee et al., *in preparation*). Gating removed low forward- and side-scatter cells before collecting a CD133-high, CD44/49b-low population. Upon collection of eight cells per slide, droplets were transferred to individual low-retention PCR tubes (eight tubes per strip) (Bioplastics K69901, B57801) pre-cooled on ice to minimize evaporation. The process was repeated with a fresh piece of parafilm for up to 12 rounds to collect 96 cells. cDNA was prepared and amplified using the SMART-Seq V4 Ultra Low Input RNA Kit (Takara Bio 634891) in 10x reduced volume reactions using a Mantis liquid handling system (S. Lee et al.*, in preparation*). Samples were stored at -20°C until quantitation and library preparation. All RNA/cDNA volumes during FACS and processing were handled using low retention siliconized pipette tips (VWR 53503-800, 535093-794).

C1 isolation of single cells and cDNA synthesis

Dissociated cells were FACS isolated as above and collected into a single 1.5 ml tube. Cells were centrifuged and resuspended at 400-800 cells/μl, combined with C1 suspension reagent, and loaded onto the C1 chip and imaged at each site to define cell number in each well. cDNA

was synthesized using SMARTer chemistry (SMARTer Ultra Low RNA Kit for the Fluidigm C1 System, Clontech 634835/634935) (three retinae, Seq 1, FW17 and FW13-1), or SMART-Seq V4 (all others). Final cDNA was harvested into low-retention tube strips and stored at -20°C.

Quality control, library preparation and sequencing

DNA was quantitated using Quant-iT PicoGreen dsDNA assay (Life Technologies, P11496) and quality checked by Bioanalyzer. Libraries > 0.05 ng/µl were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1096) and sequenced on the Illumina NextSeq 500 (2x75) (Seq 1) or on Illumina HiSeq 4000 (2x75) (all others).

5' and 3' RACE and nanopore sequencing

1 µl of amplified single cell cDNAs libraries produced as above (23 ER cells, 21 LM cells, and 5 LR cells) were re-amplified using the 5' PCR Primer II A (described for Clontech SmartSeq V4 and purchased from Integrated DNA Technologies) and using CloneAmp HiFi PCR Premix (Clontech Laboratories, Inc. Takara Bio #639298) (Table S9) by heating to 98°C x 2 min followed by 20 cycles of 98°C x 10 s, 65 °C x 15 s, and 72 °C x 35 s, followed by 72 °C x 5 min. Re-amplified cDNAs were quantified using Qubit™ Flex Fluorometer (InVitrogen) and 50 pg of each re-amplified library was used for separate RACE PCR reactions with gene-specific *NRL, RXRγ, TRβ1 and TRβ2* primers and universal 5' RACE and 3'RACE primers (Table S9) using CloneAmp HiFi PCR Premix by heating to 98°C x 2 min followed by 30 cycles of 98°C x 10 s, 60 °C x 15 s, and 72 °C x 35 s, followed by 72 °C x 5 min. The "5' RACE primer for SmartSeq cDNA" sequence (5'–AAGCAGTGGTATCAACGCAGAGTACGGG–3') was based on the SMART-Seq V4 SMARTer II A oligonucleotide (5'–AAGCAGTGGTATCAACGCAGAGTACXXXXX–3') after library sequencing revealed that AAGCAGTGGTATCAACGCAGAGTAC was most often followed by GGGNN. The "3' RACE primer for SmartSeq cDNA" sequence (5'- AAGCAGTGGTATCAACGCAGAGTACTTTT-3' was based on the SMART-Seq V4 3' SMART CDS Primer II A (5'–

AAGCAGTGGTATCAACGCAGAGTACT(30)(N-1)N–3') but with only four Ts at the 3' end. Equal volumes of RACE PCR products were pooled according to cell type, and DNA concentrations of LM (32.8 ng/μl), ER (22.6 ng/μl), and LR (16.5ng/μl) samples determined using Qubit fluorometer. Long read sequencing libraries were prepared and samples barcoded using Oxford Nanopore Technology (ONT) Native Barcoding Kit 24 V14 as specified by the manufacturer for average DNA fragment lengths of ~ 2kb. 200-300ng of each DNA pool was subjected to repair/dA-tailing (End-Prep) with NEB Ultra II End-Prep Enzyme Mix (New England Biolabs, Ipswich, MA) followed by Native Barcode Ligation as specified by ONT with NB01 assigned to the ER pool, NB02 assigned to the LM pool, and NB03 assigned to the LR pool. Ligation reactions were quenched by addition of 4 μl EDTA and each reaction mixture (24 μl) pooled for a 72 μl volume. The final pool was incubated with Native Adapter (NA) and NEB Quick T4 ligase. The adapter-ligated pool was recovered by AMPure XP magnetic beads and then loaded onto a PromethION flow cell (R10.4.1) as specified by the manufacturer. Sequencing was performed on a PromethION 24 system using MinKNOW software for 72 hours. Raw FAST5 data files were converted to FASTQ files using ONT MinKNOW base calling software.

Computational Methods

Software packages used in this study are described in Table S10.

Read processing and dimensionality reduction

Adapter sequences were removed using the *trimgalore* wrapper for *Cutadapt*[61]. Trimmed FASTQ files were used as input for *HISAT2*[62] with a non-canonical splice penalty of 20, maximum and minimum penalties for mismatch of 0 and 1, and maximum and minimum penalties for soft-clipping set to 3 and 1. Aligned bam files were quantified with *StringTie*[63] and *Tximport*[64], yielding cell-by-transcript and cell-by-gene count matrices annotated according to

Ensembl build 87. All analyses can be reproduced using a custom snakemake[65] workflow accessible at https://github.com/whtns/ARMOR.

For analysis of short-read sequencing of full-length scRNA-seq, dimensionality reduction and data visualization were performed using the Seurat (V3) toolset[66,67]. Expression counts from all sequencing datasets were integrated using Seurat's standard integration workflow. Seven sample sequencing batches were integrated after default normalization and scaling using the top 2000 most variable features in each set to identify anchor features. Normalized read counts for gene expression are reported as raw feature read counts divided by total cell read counts, then multiplied by a scaling factor of 10,000 and natural log transformed. These features of the integrated dataset were used to calculate principal component analysis (PCA) and UMAP embeddings[68,69]. A nearest-neighbors graph was constructed from the PCA embedding and clusters were identified using a Louvain algorithm at low and high resolutions (0.4 and 1.6)[70]. Read coverage plots for genes of interest were generated using *wiggleplotr*[71] to visualize BigWig files with ENSEMBL isoforms. Individual exon counts were identified using *DEXseq*[72], which takes exons from available isoforms and bins them into intervals before assigning any read counts that overlap a bin to that same bin. These counts were normalized for length for calculation of fold change or relative difference of exon use. Reads crossing target splice donor sites were evaluated as spliced or unspliced from BAM files to determine intron readthrough.

For analysis of 3' snRNA-seq, processed data provided with Zuo *et al*[20] (CZ CELLxGENE Discover with accession code 5900dda8-2dc3-4770- b604084eac1c2c82) was downloaded as a Seurat Object. A pre-processed AnnData object containing inferred NRPC fate annotations was generously provided by Dr. Rui Chen, cone and rod fated NRPC annotations were transferred to the primary Seurat Object, and data visualization was performed using the Seurat (V5) toolset[66,67].

Analysis of long-read sequencing data was performed with the nf-core nanoseq version 3.1.0 pipeline[73] (https://nf-co.re/nanoseq/3.1.0). Quality control on raw reads was conducted using FastQC. Reads were aligned using minimap2[74]. SAM files were converted to coordinate-sorted BAM files and mapping metrics were obtained using SAMtools. bigWig coverage tracks were created for visualization using BEDTools and bedGraphToBigWig. bigBed coverage tracks were created using BEDTools and bedToBigBed. Transcripts were reconstructed and quantified using bambu. Quality control results for raw reads and alignment results were presented using MultiQC. All reads whose ends overlapped gene-specific RACE primers were displayed and quantitated using the Integrated Genome Browser (IGV). For *THRB* 3' RACE reactions, reads inferred to initiate through exonic internal oligo(dT) priming were removed based on the following features adjacent to the 3′ end of the alignments: six continuous adenines (As), more than seven As in a 10 nucleotide window, AG-runs of six or more nucleotides, eight or more As or Gs in a 10 nucleotide window, eight or more As or high A/T content (27 out of 30 bases), or 12 or more adenines present in an 18 nt window[75]. The proportion of alternative *NRL* exon 2 splicing was calculated based on the maximum read coverage between the exon 2 splice acceptor and the first splice donor and the minimum coverage after the internal splice donor as reported in IGV.

Differential expression and overrepresentation analyses

Marker features for each cluster were identified using the Wilcoxon-rank sum tests and specificity scores were computed using the *genesorteR*[76] R package. Only genes with an adjusted p-value (pAdj) <0.05 were considered for further analyses. Differential expression analysis of full-length scRNA-seq was performed using the Seurat FindMarkers function with the Wilcoxon-rank sum test on log-normalized counts. Results were displayed as volcano plots made with *EnhancedVolcano*[77]. Differential expression analysis of a 3' snRNA-seq was

performed using the Seurat FindMarkers function with the Wilcoxon-rank sum test on log-normalized counts. Results were displayed as volcano plots made with *EnhancedVolcano*[77].

Overrepresentation analyses were performed with WebGestalt[78]. Analyses were run with the default settings (5-2000 genes per category, Benjamini-Hochberg multiple-testing correction), gene list enrichment was compared to the genome reference set and ontologies were displayed with a false discovery rate (FDR) <0.05 and 'weighted set cover' redundancy reduction. All gene sets evaluated were provided within WebGestalt and the same three were used unless otherwise indicated: GO – Biological Process, KEGG, and Hallmark50.

Transcription factor regulons

Transcription factor regulons were identified using pySCENIC, the python-based implementation of Single-Cell Regulatory Network Inference and Clustering (SCENIC)[30,79]. The tool was run using the basic settings as shown (https://pyscenic.readthedocs.io/en/latest/tutorial.html). Initial filtering required a gene to have a minimum of 3 raw counts in 1% of cells to be considered for inclusion in a regulon. AUC scores were z-score normalized for comparison between regulons.

Trajectory analysis

RNA velocity analysis was performed with R package *velocyto*[39]. Spliced and unspliced count matrices were assembled using the run_smartseq2 subcommand with repeat-masked ENSEMBL build 87. RNA velocity was calculated using cell kNN pooling with a gamma fit based on a 95% quantile and velocities were overlaid on integrated UMAP visualizations using *velocyto*. Pseudotemporal trajectories were calculated using Monocle 3[80]. The SeuratWrappers package provided integration between Seurat's final integrated dimensional reduction and Monocle. The data was subset to include only cone-directed iCPs and L/M cones, then a principal graph was fit across the UMAP plot and a 'root cell' chosen to represent the latest maturation point in the *OPN1LW*+ late-maturing cone group. After assigning pseudotime values to each cell, genes that significantly changed as a function of pseudotime were identified and

those with a correlation q-value of <0.05 with expression > 0.5 in at least 5% of the cells present in the pseudotime were grouped into modules of co-regulation by performing UMAP and Louvain community analysis.

Histology

Fixation and cryosectioning of fetal retina

Retinae were procured and dissected in cold 1x PBS. The cornea and lens were removed with a cut around the limbus of the iris leaving the front of the eye open. The tissue was submerged in ~25 ml of 4% paraformaldehyde (PFA) in 1x TBS and placed at 4°C on a rocker at low speed for ~16-18 hrs. The tissue was washed 3 times with 1x PBS before incubation in 30% sucrose [sucrose + 1x PBS] overnight at 4°C. A mold was made by cutting off the top 5 ml of a 50 ml conical tube and placed on dry ice. A solution of 1:2 OCT Compound:30% sucrose was mixed and centrifuged to remove bubbles. The mold was partially filled before adding the dissected eye and then covering fully. 10 µm frozen sections were cut at ~-24°C with the front of the eye facing side on to the blade. Explanted retinae were fixed for 15 min in 4% PFA, washed with 1x PBS, incubated in 30% sucrose for 15 min, after which the membrane was cut from its frame and transferred to a mold containing 1:2 OCT:30% sucrose and frozen.

Immunohistochemical staining

For SYK and ARR3 co-staining, tissue sections were warmed at RT until dry and then washed in Coplin jars twice with 1x PBS for 5 min at RT. Samples were permeabilized in 1x PBS-T for 5 min, washed again in 1x PBS twice for 5 min, blocked for 1 hr in blocking solution [1x PBS, 0.1% Triton X-100, 5% normal donkey serum, 5% normal horse serum]. Primary mouse anti-SYK and rabbit anti-ARR3 (LUMIF-hCAR) antibodies were mixed in blocking solution and applied to samples overnight (16-18 h) at 4°C. Slides were washed three times with 1x PBS for 10 min each, probed with secondary antibodies (Key Resources Table) in blocking solution for 1

hr at RT, washed twice 1x PBS for 15 min each, incubated in1x PBS containing 10 µg/ml DAPI for 10 min, and mounted with Mowiol solution. Additional immunohistological staining of tissue sections and cells on coverslips was performed as described[11] but without initial EDTA wash.

Combined in situ hybridization and immunohistological staining

Human retinal sections were prepared as above but using RNase-free reagents for all stages (PBS, TBS) and ultrapure sucrose (JT Baker, 4097-04) and cryostats cleaned with ELIMINase (Decon Labs, Inc) and 70% ethanol prior to use. RNA FISH hybridization chain reaction probes were designed by and obtained from Molecular Instruments, Inc, with 20 20-bp single strand DNA probes per target RNA sequence except Tr-NRL exon 1 (4 probes), FL-NRL exons 1 and (8 probes), and *HOTAIRM1* (14 probes). The accession numbers for target sequences are shown in Table S11, except for *RP13-143G15.4* which included regions of several isoforms for maximum gene coverage. The *in situ* protocol was performed with fluorescent hairpins and buffers from Molecular Instruments using the manufacturer's instructions for mammalian cells on a slide (https://files.molecularinstruments.com/MI-Protocol-RNAFISH-MammalianCellSlide-Rev7.pdf), beginning with a 4% PFA treatment after thawing and drying slides at RT for 10 min, with the following changes: samples were incubated in 70% ethanol overnight in a Coplin jar, probes were added to the final hybridization volume at 4x the recommended concentration for tissue sections (2.4 pmol in 150 µl of hybridization buffer), tissues were pre-hybridized at 42°C and then lowered to 37°C once probes were added, and hairpins were used at a volume of 1 µl (6 pmol) in 50 µl amplification buffer per section. After removing amplification solution and washing, samples were directly used in the immunofluorescence protocol. Briefly, samples were blocked in a basic RNase-free solution (1% BSA, 0.1% Triton X-100, 0.05% Tween-20 in PBS) for 1 hr and primary antibodies mixed in blocking solution and applied to samples overnight at 4°C. Samples were washed three times for 10 min with 1x PBS, then incubated with secondary antibodies in blocking solution for 40 min at RT. Samples were washed as above with DAPI in

the third and final wash, and mounted with Mowiol before imaging using a Leica STELLARIS 5 inverted confocal microscope. For *GNAT2* and *NR2E3* FISH, samples were washed with DAPI in the third and final wash, mounted and then imaged as above. After this first imaging, coverslips were removed, and slides were washed in PBS and then used in the immunofluorescence protocol with an alternate blocking solution) (2.5% Horse, donkey, human sera, 1% BSA, 0.1% Triton X-100, 0.05% Tween-20 in TBS).

Quantitation of FISH

For *NRL*, *CHRNA1*, and lncRNAs, QuPath[81] was used to identify nuclei, predict cell expansion, classify cells, and count FISH puncta. Apical retina regions containing RXRγ+ and/or NRL+ cells were visually selected for evaluation. The StarDist extension[82] was used to outline nuclei and cell expansion zones with deep learning model '*dsb2018_heavy_augment.pb'* on the DAPI image channel (threshold = 0.6, pixelSize = 0.15). Model results were reviewed and cells with partial, multiple, or overlapping nuclei removed from consideration. FISH puncta were thresholded then output, using the subcellular detection function (expected spot size = 0.5 μm^2, min spot size = 0.2 μm^2, max spot size = 0.7 μm^2 (*NRL* isoforms) or 1 μm^2 (*CHRNA1* and lncRNAs) to capture the most puncta above background. Image positions were determined by tracing the retina apical edge with line segments in FIJI, identifying the closest point on the line to the midpoint of each image, then measuring the distance from the line starting position. For *GNAT2* and *NR2E3* FISH, images were analyzed in FIJI. The outer layer of cells was delimited manually based on the RXRγ+ cells and the number of RXRγ+ and/or *GNAT2*+ and/or *NR2E3*+ cells were manually counted, examining each color channel separately. For the inner layer, the number of *NR2E3* RNA positive cells and NR2E3 protein positive cells were counted.

<u>*NRL* isoform analyses</u>

Proteasome inhibition of CHLA-VC-RB-31

10^6 CHLA-VC-RB-31 cells[37] were cultured in 2 ml retina culture medium in a 24-well plate with or without 10 μM retinoic acid (Cayman Chemical 11017) and incubated at 37°C in 5% CO_2. MG132 (Sigma-Aldrich 474790) was added at 23 h to a final concentration of 10 μM and cells incubated for 4 h. Cells were collected and centrifuged at 300 x g for 4 min, resuspended in 60 μl cold 1x RIPA buffer (Millipore 20-188), 10% sodium dodecyl sulfate (SDS), protease and phosphatase inhibitors diluted per manufacturer instructions (Sigma-Aldrich 5892970001, 4906837001)], incubated on ice for 30 min, then centrifuged at 4°C for 20 min at 14,000 RPM. 15 μl of supernatant were mixed with 3 μl 6X sample buffer (300 mM Tris, 60% glycerol,12% SDS, 86 mM B-mercaptoethanol, 0.6 mg/ml bromophenol blue), heated to 95°C for 5 min, and separated on a 4-12% Bis-Tris gel (Invitrogen, NP0321) using 1xMOPS running buffer (Life Technologies NP0001). Protein was wet transferred to a PVDF membrane in Towbin transfer buffer (25 mM Tris, 192 mM glycine pH 8.3,10% methanol) at 20 V ON at RT. The membrane was blocked in 1xTBS-T containing 5% w/v nonfat dry milk for 1 hr, washed with TBS-T three times for 5 min, incubated overnight in 0.1% dry milk TBS-T solution containing primary antibody at 4°C with slow rocking, washed three times for 15 min in TBS-T, incubated with HRP-conjugated secondary antibody for 1 hr at RT in the 0.1% milk solution, washed six times for 15 min, incubated in chemiluminescent substrate (Thermo Fisher, 34094) and imaged.

NRL isoform and reporter constructs

The pcDNA4-His-Max-C-Nrl[83] (gift of A. Swaroop) CMV promoter was replaced with EF1α promoter to produce pcDNA4-His-Max-C-EF1α-FL-NRL. Using In-Fusion (Takara Bio) the His-Max tag was removed to produce full-length NRL expression plasmid pCDNA4-C-EF1α-FL-NRL (primers #1,2). The His-Max tag plus first 417 bp of NRL N-terminal coding sequence were removed to produce truncated NRL expression plasmid pCDNA4-C-EF1α-Tr-NRL (primers #1,2,3,4). pcDNA4-C-EF1α-empty transfection carrier plasmid was made by removing NRL coding sequence from pCDNA4-C-EF1α-Fl-NRL (primers #5,6). pGL3-PDE6B-146-luc was

produced by amplifying a 146 bp promoter region of *PDE6B*[84] from H9 hESC genomic DNA and inserting into a pGL3-SV40 (Promega, primers #7,8), then removing the SV40 promoter (primers #9,10) to generate pGL3-PDE6B-146. The promoter-less pGL3-empty was produced by removing SV40 sequences from pGL3-SV40 (primers #9,11). pUltra-EGFP-P2A-Tr-NRL was produced by amplification of NRL coding sequence from pcDNA4-C-EF1α-FL-NRL and inserting into BamHI digested pUltra-EGFP (Addgene 24129) (primers #12, 13). All primers are listed in the Key Resources Table.

NRL expression and analysis

pcDNA4-C-EF1α-NRL constructs were transfected into HEK293T cells. 4 µg plasmid in 1 ml DMEM (Life Technologies, 10313039) was combined with 1 ml DMEM + 12 µg polyethylenimine (PEI MAX) (MW 40,000, Polysciences 24765-1) and combined with 10^6 trypsinized HEK293T cells, plated into two wells of a 6-well plate with an additional 1 ml of DMEM + 10% FBS, then cultured overnight in a 37°C incubator. Media was replaced with fresh 293T media the next day and after two additional days cells were collected, lysed, and protein quantitated via BCA assay (Thermo Scientific PI23227). 30 µg of each well was used for western blot as above, with membranes probed with antibody raised against the NRL N-terminus (Santa Cruz SC-374277) or against full-length NRL (R&D AF2945).

Luciferase assays

575 ng of total DNA (50 ng reporter, 100ng FL-NRL and/or 100-300 ng Tr-NRL expression vector, and the remainder pcDNA4-C-EF1α empty vector) in 5 µl DMEM was mixed with 2.3 µg PEI in another 5 µl DMEM (per condition, in triplicate) and then plated into one well of a white bottom 96-well cell culture plate with 50,000 NIH3T3 cells in 100 µl 293T media. The luciferase assay was carried out 48 h after transfection with the Nano-Glo Dual-Luciferase Reporter Assay

System (Promega N1610) per manufacturer instructions. Luminescence was recorded using the Promega Glomax Multi+ Detection System.

Truncated NRL overexpression in fetal retina

Concentrated pUltra-EGFP-NRL-205 and control pUltra-EGFP lentivirus were produced as described[85]. At 16-24 h after 293T transfection in 15 cm dishes, media was exchanged for 20 ml UltraCULTURE (Lonza Inc., BP12-725F) containing 1% HEPES (Sigma, 0887), 1% GlutaMax (Life Technologies 35050061), and 1% Penicillin-Streptomycin (Fisher Scientific, MT30002CI). After 60-64 h, media was harvested, centrifuged at 3000 x g for 10 min at 4°C, filtered through 0.45 µm PVDF flask filter, concentrated via tangential flow filtration using a MidiKros 20 cm 500 KD 0.5 mm column (D02-5500-05-S), with both input and final supernatant containers kept in ice. The concentrate was re-filtered using a 0.45 µm PVDF syringe filter (Millipore), stored in aliquots at -80°C, and titered using a p24 ELISA kit (ZeptoMetrix 801002).

As described[11], fetal eyes were washed in 70% ethanol, washed three times in 1x sterile PBS, and dissected in cold 1x PBS. The cornea was removed and tissue cut ~one-third of the distance towards the posterior pole along 4 lines approximately equidistant and between attached tendons. The retina was removed in this flattened state after cutting the optic nerve, transferred to a polytetrafluoroethylene culture insert (Millipore, PICM0RG50) with photoreceptor side down, and placed in a 6-well plate with 1,200 µl retinal culture medium. Retinae were infected as described, cultured for 7 days with half media changes every other day, then fixed in 4% PFA for 15 min, washed three times in 1x PBS, equilibrated in sucrose (30% in PBS) for 15 min, and embedded and sectioned as above. Immunostaining and imaging were performed as above.

<u>*RB1* knockdown and SYK inhibitor treatment</u>

FW18.5 retina was partially dissociated with papain, cultured overnight in retinal culture media at 37°C in a 6-well plate, then frozen in 10% DMSO solution and stored in liquid nitrogen. Samples were thawed and revived in retinal culture media overnight and infected with lentivirus carrying shRNAs targeting *RB1* (sh*RB1-733*)[10] or a scrambled control (sh*SCR*) and diluted β-mercaptoethanol, insulin, glutamine, penicillin, and streptomycin to the final retina culture medium concentrations along with 5 µg/ml Polybrene, for 24 hrs. After replacing 2/3 of media in each well with fresh media, cells were treated with GS-9876 (MedChemExpress, HY-109091) in DMSO at or with the same volume of DMSO vehicle control. After 12 d, cells were attached to poly-L-Lysine-coated coverslips for 3 h, fixed in 4% PFA for 10 min, washed in 1x PBS three times, and stored at -20°C until immunocytological staining with antibodies to Ki67, RXRγ, and GFP (for YFP staining) as described[10].

Statistical Analysis

Statistical methods and packages are cited in the Methods text and figure legends. Statistical comparison between more than two groups of cells for gene expression, regulon activity, or exon count proportions used the non-parametric Kruskal-Wallis test followed by pairwise post-hoc Dunn tests with Benjamini-Hochberg correction. Mean expression levels of *NRL* and *RXRG* isoforms across cell clusters were evaluated by Welch's t-tests for groups with unequal variance; p-values for these tests were estimated using bootstrapping with 1,000,000 replications per comparison. Changes in FISH puncta counts were examined within each channel individually using zero-inflated negative binomial regression models with robust standard error terms. Distance across the sample tissue was used to model degenerate zero inflation. P values reported in the manuscript text and legends are Wald Chi-Squared tests that evaluate differences in count between inflection points. Analyses were performed using Microsoft Excel, the R statistical language in RStudio, or Stata SE v14.2.

Acknowledgements

We thank Melissa L. Wilson (USC Department of Preventive Medicine) and Family Planning Associates for assistance in obtaining fetal tissue, Tingting Yang and Rui Chen for sharing data, the Stem Cell Analytics Core Facility, FACS Core Facility, and Imaging Core Facility of The Saban Research Institute of Children's Hospital Los Angeles for technical support, and Biraj Mahato and Sumitha Bharathan for critical reading an earlier version of the manuscript.

Funding:

National Institutes of Health grant R01EY026661 (DC)
National Institutes of Health grant R01CA137124 (DC)
National Institutes of Health grant R01AG076956 (MAB)
National Institutes of Health grant 5T32HD060549 to USC Department of Development, Stem Cells, and Regenerative Medicine (DWHS)
USC Provost Fellowship (DWHS)
Saban Research Institute of Children's Hospital Los Angeles fellowship (JB)
Research to Prevent Blindness (unrestricted grant to USC Dept. of Ophthalmology)
Larry and Celia Moh Foundation (DC)
Neonatal Blindness Research Fund (DC)
A.B. Reins Foundation (DC)
Knights Templar Eye Foundation (DC)

Author contributions:

Conceptualization: DWHS, DC
Methodology: DWHS, LC, DJW, HPS, DC
Investigation: DWHS, LC, SL, JB, DJW, JGA, YK, HPS, ZF
Computational analyses: KS, BB, MS
Statistical analyses: DWHS, MWR
Bioinformatic approaches: KS, MB, MAB
Computational infrastructure: ED, SGE
Tissue procurement: MT, BHG
Key reagent: CMC
Writing --- original draft: DWHS, DC
Writing --- review and editing: DWHS, KS, JGA, MAB, BHG, CMC, DC

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. A web portal providing access to a Shiny app for manipulation of the presented full-length scRNA-seq data (including UMAP or tSNE display of clusters, gene or isoform expression, and sample metadata; cluster marker gene and violin plots; exon coverage plots; Ensembl isoform assignments) is available at https://docker.saban.chla.usc.edu/cobrinik/app/seuratApp/. The preprocessed Seurat Object for

full-length scRNAseq (and for the above-linked Shiny app) is available at https://doi.org/10.5281/zenodo.15231490. The Seurat Object for 3' snRNA-seq[20] was downloaded from CZ CELLxGENE Discover with accession code 5900dda8-2dc3-4770-b604084eac1c2c82. Single-cell RNA-sequencing files and processed final gene and transcript matrices are publicly available at the GEO database under accession number GSE207802.

References

1. Shiau, F., Ruzycki, P.A., and Clark, B.S. (2021). A single-cell guide to retinal development: Cell fate decisions of multipotent retinal progenitors in scRNA-seq. Dev Biol *478*, 41–58. https://doi.org/10.1016/j.ydbio.2021.06.005.

2. Brzezinski, J.A., and Reh, T.A. (2015). Photoreceptor cell fate specification in vertebrates. Development *142*, 3263–3273. https://doi.org/10.1242/dev.127043.

3. Emerson, M.M., Surzenko, N., Goetz, J.J., Trimarchi, J., and Cepko, C.L. (2013). Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors. Dev Cell *26*, 59–72. https://doi.org/10.1016/j.devcel.2013.06.005.

4. Ng, L., Lu, A., Swaroop, A., Sharlin, D.S., Swaroop, A., and Forrest, D. (2011). Two Transcription Factors Can Direct Three Photoreceptor Outcomes from Rod Precursor Cells in Mouse Retinal Development. The Journal of Neuroscience *31*, 11118–11125. https://doi.org/10.1523/JNEUROSCI.1709-11.2011.

5. Hendrickson, A., Possin, D., Vajzovic, L., and Toth, C.A. (2012). Histologic Development of the Human Fovea From Midgestation to Maturity. Am J Ophthalmol *154*, 767-778.e2. https://doi.org/10.1016/j.ajo.2012.05.007.

6. Hoshino, A., Ratnapriya, R., Brooks, M.J., Chaitankar, V., Wilken, M.S., Zhang, C., Starostik, M.R., Gieser, L., La Torre, A., Nishio, M., et al. (2017). Molecular Anatomy of the Developing Human Retina. Dev Cell *43*, 763-779.e4. https://doi.org/10.1016/j.devcel.2017.10.029.

7. Georgiou, M., Robson, A.G., Fujinami, K., de Guimarães, T.A.C., Fujinami-Yokokawa, Y., Daich Varela, M., Pontikos, N., Kalitzeos, A., Mahroo, O.A., Webster, A.R., et al. (2024). Phenotyping and genotyping inherited retinal diseases: Molecular genetics, clinical and imaging features, and therapeutics of macular dystrophies, cone and cone-rod dystrophies, rod-cone dystrophies, Leber congenital amaurosis, and cone dysfunction syndromes. Prog Retin Eye Res *100*, 101244. https://doi.org/10.1016/j.preteyeres.2024.101244.

8. Cobrinik, D. (2024). Retinoblastoma Origins and Destinations. New England Journal of Medicine *390*, 1408–1419. https://doi.org/10.1056/NEJMra1803083.

9. Xu, X.L., Fang, Y., Lee, T.C., Forrest, D., Gregory-Evans, C., Almeida, D., Liu, A., Jhanwar, S.C., Abramson, D.H., and Cobrinik, D. (2009). Retinoblastoma Has Properties of a Cone Precursor Tumor and Depends Upon Cone-Specific MDM2 Signaling. Cell *137*, 1018–1031. https://doi.org/10.1016/j.cell.2009.03.051.

10. Xu, X.L., Singh, H.P., Wang, L., Qi, D.L., Poulos, B.K., Abramson, D.H., Jhanwar, S.C., and Cobrinik, D. (2014). Rb suppresses human cone-precursor-derived retinoblastoma tumours. Nature *514*, 385–388. https://doi.org/10.1038/nature13813.

11. Singh, H.P., Wang, S., Stachelek, K., Lee, S., Reid, M.W., Thornton, M.E., Craft, C.M., Grubbs, B.H., and Cobrinik, D. (2018). Developmental stage-specific

proliferation and retinoblastoma genesis in RB-deficient human but not mouse cone precursors. Proceedings of the National Academy of Sciences *115*, E9391–E9400. https://doi.org/10.1073/pnas.1808903115.

12. McEvoy, J., Flores-Otero, J., Zhang, J., Nemeth, K., Brennan, R., Bradley, C., Krafcik, F., Rodriguez-Galindo, C., Wilson, M., Xiong, S., et al. (2011). Coexpression of Normally Incompatible Developmental Pathways in Retinoblastoma Genesis. Cancer Cell *20*, 260–275. https://doi.org/10.1016/j.ccr.2011.07.005.

13. Khanna, H., Akimoto, M., Siffroi-Fernandez, S., Friedman, J.S., Hicks, D., and Swaroop, A. (2006). Retinoic Acid Regulates the Expression of Photoreceptor Transcription Factor NRL. Journal of Biological Chemistry *281*, 27327–27334. https://doi.org/10.1074/jbc.M605500200.

14. Clark, B.S., Stein-O'Brien, G.L., Shiau, F., Cannon, G.H., Davis-Marcisak, E., Sherman, T., Santiago, C.P., Hoang, T. V., Rajaii, F., James-Esposito, R.E., et al. (2019). Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. Neuron *102*, 1111-1126.e5. https://doi.org/10.1016/j.neuron.2019.04.010.

15. Lu, Y., Shiau, F., Yi, W., Lu, S., Wu, Q., Pearson, J.D., Kallman, A., Zhong, S., Hoang, T., Zuo, Z., et al. (2020). Single-Cell Analysis of Human Retina Identifies Evolutionarily Conserved and Species-Specific Mechanisms Controlling Development. Dev Cell *53*, 473-491.e9. https://doi.org/10.1016/j.devcel.2020.04.009.

16. Sridhar, A., Hoshino, A., Finkbeiner, C.R., Chitsazan, A., Dai, L., Haugan, A.K., Eschenbacher, K.M., Jackson, D.L., Trapnell, C., Bermingham-McDonogh, O., et al. (2020). Single-Cell Transcriptomic Comparison of Human Fetal Retina, hPSC-Derived Retinal Organoids, and Long-Term Retinal Cultures. Cell Rep *30*, 1644-1659.e4. https://doi.org/10.1016/j.celrep.2020.01.007.

17. Buenaventura, D.F., Corseri, A., and Emerson, M.M. (2019). Identification of genes with enriched expression in early developing mouse cone photoreceptors. Invest Ophthalmol Vis Sci *60*, 2787–2799. https://doi.org/10.1167/iovs.19-26951.

18. Lo Giudice, Q., Leleu, M., La Manno, G., and Fabre, P.J. (2019). Single-cell transcriptional logic of cell-fate specification and axon guidance in early born retinal neurons. Development *146*. https://doi.org/10.1242/dev.178103.

19. Lyu, P., Hoang, T., Santiago, C.P., Thomas, E.D., Timms, A.E., Appel, H., Gimmen, M., Le, N., Jiang, L., Kim, D.W., et al. (2021). Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. Cell Rep *37*, 109994. https://doi.org/10.1016/j.celrep.2021.109994.

20. Zuo, Z., Cheng, X., Ferdous, S., Shao, J., Li, J., Bao, Y., Li, J., Lu, J., Jacobo Lopez, A., Wohlschlegel, J., et al. (2024). Single cell dual-omic atlas of the human developing retina. Nat Commun *15*, 1–21. https://doi.org/10.1038/s41467-024-50853-5.

21. Omori, Y., Kubo, S., Kon, T., Furuhashi, M., Narita, H., Kominami, T., Ueno, A., Tsutsumi, R., Chaya, T., Yamamoto, H., et al. (2017). Samd7 is a cell type-specific PRC1 component essential for establishing retinal rod photoreceptor identity. Proceedings of the National Academy of Sciences *114*, E8264–E8273. https://doi.org/10.1073/pnas.1707021114.

22. Gagliardi, G., Ben M'Barek, K., Chaffiol, A., Slembrouck-Brec, A., Conart, J.-B., Nanteau, C., Rabesandratana, O., Sahel, J.-A., Duebel, J., Orieux, G., et al. (2018). Characterization and Transplantation of CD73-Positive Photoreceptors Isolated from Human iPSC-Derived Retinal Organoids. Stem Cell Reports *11*, 665–680. https://doi.org/10.1016/j.stemcr.2018.07.005.

23. Chen, T.W., Chen, G., Funkhouser, L.J., and Nam, S.-C. (2009). Membrane domain modulation by Spectrins in Drosophila photoreceptor morphogenesis. genesis *47*, 744–750. https://doi.org/10.1002/dvg.20555.

24. Burger, C.A., Jiang, D., Mackin, R.D., and Samuel, M.A. (2021). Development and maintenance of vision's first synapse. Dev Biol *476*, 218–239. https://doi.org/10.1016/j.ydbio.2021.04.001.

25. Sokolov, D., Sechrest, E.R., Wang, Y., Nevin, C., Du, J., and Kolandaivelu, S. (2021). Nuclear NAD+-biosynthetic enzyme NMNAT1 facilitates development and early survival of retinal neurons. Elife *10*, 1–26. https://doi.org/10.7554/eLife.71185.

26. Lukowski, S.W., Lo, C.Y., Sharov, A.A., Nguyen, Q., Fang, L., Hung, S.S., Zhu, L., Zhang, T., Grünert, U., Nguyen, T., et al. (2019). A single-cell transcriptome atlas of the adult human retina. EMBO J *38*, 1–15. https://doi.org/10.15252/embj.2018100811.

27. Peng, Y.R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S., van Zyl, T., Do, M.T.H., Regev, A., and Sanes, J.R. (2019). Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. Cell *176*, 1222-1237.e22. https://doi.org/10.1016/j.cell.2019.01.004.

28. Kallman, A., Capowski, E.E., Wang, J., Kaushik, A.M., Jansen, A.D., Edwards, K.L., Chen, L., Berlinicke, C.A., Joseph Phillips, M., Pierce, E.A., et al. (2020). Investigating cone photoreceptor development using patient-derived NRL null retinal organoids. Commun Biol *3*, 82. https://doi.org/10.1038/s42003-020-0808-5.

29. Voigt, A.P., Whitmore, S.S., Flamme-Wiese, M.J., Riker, M.J., Wiley, L.A., Tucker, B.A., Stone, E.M., Mullins, R.F., and Scheetz, T.E. (2019). Molecular characterization of foveal versus peripheral human retina by single-cell RNA sequencing. Exp Eye Res *184*, 234–242. https://doi.org/10.1016/j.exer.2019.05.001.

30. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc *15*, 2247–2276. https://doi.org/10.1038/s41596-020-0336-2.

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

32. Ng, L., Hurley, J.B., Dierks, B., Srinivas, M., Saltó, C., Vennström, B., Reh, T.A., and Forrest, D. (2001). A thyroid hormone receptor that is required for the development of green cone photoreceptors. Nat Genet *27*, 94–98. https://doi.org/10.1038/83829.

33. Roberts, M.R., Hendrickson, A., McGuire, C.R., and Reh, T.A. (2005). Retinoid X Receptor γ Is Necessary to Establish the S-opsin Gradient in Cone Photoreceptors of the Developing Mouse Retina. Investigative Opthalmology & Visual Science *46*, 2897. https://doi.org/10.1167/iovs.05-0093.

34. Friedman, J.S., Khanna, H., Swain, P.K., DeNicola, R., Cheng, H., Mitton, K.P., Weber, C.H., Hicks, D., and Swaroop, A. (2004). The Minimal Transactivation Domain of the Basic Motif-Leucine Zipper Transcription Factor NRL Interacts with TATA-binding Protein. Journal of Biological Chemistry *279*, 47233–47241. https://doi.org/10.1074/jbc.M408298200.

35. Rehemtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D.J., and Swaroop, A. (1996). The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. Proceedings of the National Academy of Sciences *93*, 191–195. https://doi.org/10.1073/pnas.93.1.191.

36. Swaroop, A., Xu, J.Z., Pawar, H., Jackson, A., Skolnick, C., and Agarwal, N. (1992). A conserved retina-specific gene encodes a basic motif/leucine zipper domain. Proceedings of the National Academy of Sciences *89*, 266–270. https://doi.org/10.1073/pnas.89.1.266.

37. Stachelek, K., Harutyunyan, N., Lee, S., Beck, A., Kim, J., Xu, L., Berry, J.L., Nagiel, A., Reynolds, C.P., Murphree, A.L., et al. (2023). Non-synonymous, synonymous, and non-coding nucleotide variants contribute to recurrently altered biological processes during retinoblastoma progression. Genes Chromosomes Cancer *62*, 275–289. https://doi.org/10.1002/gcc.23120.

38. Eldred, K.C., Hadyniak, S.E., Hussey, K.A., Brenerman, B., Zhang, P.-W., Chamling, X., Sluch, V.M., Welsbie, D.S., Hattar, S., Taylor, J., et al. (2018). Thyroid hormone signaling specifies cone subtypes in human retinal organoids. Science (1979) *362*, eaau6348. https://doi.org/10.1126/science.aau6348.

39. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

40. Roesch, K., Jadhav, A.P., Trimarchi, J.M., Stadler, M.B., Roska, B., Sun, B.B., and Cepko, C.L. (2008). The transcriptome of retinal Müller glial cells. Journal of Comparative Neurology *509*, 225–238. https://doi.org/10.1002/cne.21730.

41. Pereiro, X., Miltner, A.M., La Torre, A., and Vecino, E. (2020). Effects of Adult Müller Cells and Their Conditioned Media on the Survival of Stem Cell-Derived Retinal Ganglion Cells. Cells *9*, 1759. https://doi.org/10.3390/cells9081759.

42. Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W.P., Weber, G., Lee, K., Fraioli, R.E., Cho, S.-H., et al. (2004). Genomic Analysis of Mouse Retinal Development. PLoS Biol *2*, e247. https://doi.org/10.1371/journal.pbio.0020247.

43. Walcott, J.C., and Provis, J.M. (2003). Müller cells express the neuronal progenitor cell marker nestin in both differentiated and undifferentiated human foetal retina. Clin Exp Ophthalmol *31*, 246–249. https://doi.org/10.1046/j.1442-9071.2003.00638.x.

44. Finkbeiner, C., Ortuño-Lizarán, I., Sridhar, A., Hooper, M., Petter, S., and Reh, T.A. (2022). Single-cell ATAC-seq of fetal human retina and stem-cell-derived retinal organoids shows changing chromatin landscapes during cell fate acquisition. Cell Rep *38*. https://doi.org/10.1016/j.celrep.2021.110294.

45. Marquardt, T., Ashery-Padan, R., Andrejewski, N., Scardigli, R., Guillemot, F., and Gruss, P. (2001). Pax6 Is Required for the Multipotent State of Retinal Progenitor Cells. Cell *105*, 43–55. https://doi.org/10.1016/S0092-8674(01)00295-1.

46. Poché, R.A., Furuta, Y., Chaboissier, M.C., Schedl, A., and Behringer, R.R. (2008). Sox9 is expressed in mouse multipotent retinal progenitor cells and functions in Müller Glial cell development. Journal of Comparative Neurology *510*, 237–250. https://doi.org/10.1002/cne.21746.

47. Hafler, B.P., Surzenko, N., Beier, K.T., Punzo, C., Trimarchi, J.M., Kong, J.H., and Cepko, C.L. (2012). Transcription factor Olig2 defines subpopulations of retinal progenitor cells biased toward specific cell fates. Proceedings of the National Academy of Sciences *109*, 7882–7887. https://doi.org/10.1073/pnas.1203138109.

48. Zhang, J., Benavente, C.A., McEvoy, J., Flores-Otero, J., Ding, L., Chen, X., Ulyanov, A., Wu, G., Wilson, M., Wang, J., et al. (2012). A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature *481*, 329–334. https://doi.org/10.1038/nature10733.

49. Blomgren, P., Chandrasekhar, J., Di Paolo, J.A., Fung, W., Geng, G., Ip, C., Jones, R., Kropf, J.E., Lansdon, E.B., Lee, S., et al. (2020). Discovery of Lanraplenib (GS-9876): A Once-Daily Spleen Tyrosine Kinase Inhibitor for Autoimmune Diseases. ACS Med Chem Lett *11*, 506–513. https://doi.org/10.1021/acsmedchemlett.9b00621.

50. Kamieniarz-Gdula, K., and Proudfoot, N.J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. Trends in Genetics *35*, 553–564. https://doi.org/10.1016/j.tig.2019.05.005.

51. Master, A., and Nauman, A. (2014). THRB (Thyroid Hormone Receptor, Beta). Atlas Genet Cytogenet Oncol Haematol *18*, 400–433. https://doi.org/10.4267/2042/53769.

52. Mercer, T.R., Wilhelm, D., Dinger, M.E., Soldà, G., Korbie, D.J., Glazov, E.A., Truong, V., Schwenke, M., Simons, C., Matthaei, K.I., et al. (2011). Expression of distinct RNAs from 3′ untranslated regions. Nucleic Acids Res *39*, 2393–2403. https://doi.org/10.1093/nar/gkq1158.

53. Fischer, A.J., Foster, S., Scott, M.A., and Sherwood, P. (2008). Transient expression of LIM-domain transcription factors is coincident with delayed maturation of photoreceptors in the chicken retina. J Comp Neurol *506*, 584–603. https://doi.org/10.1002/cne.21578.

54. Irie, S., Sanuki, R., Muranishi, Y., Kato, K., Chaya, T., and Furukawa, T. (2015). Rax Homeoprotein Regulates Photoreceptor Cell Maturation and Survival in Association with Crx in the Postnatal Mouse Retina. Mol Cell Biol *35*, 2583–2596. https://doi.org/10.1128/MCB.00048-15.

55. Chen, C.-M.A., and Cepko, C.L. (2002). The chicken RaxL gene plays a role in the initiation of photoreceptor differentiation. Development *129*, 5363–5375. https://doi.org/10.1242/dev.00114.

56. Welby, E., Lakowski, J., Di Foggia, V., Budinger, D., Gonzalez-Cordero, A., Lun, A.T.L., Epstein, M., Patel, A., Cuevas, E., Kruczek, K., et al. (2017). Isolation and Comparative Transcriptome Analysis of Human Fetal and iPSC-Derived Cone Photoreceptor Cells. Stem Cell Reports *9*, 1898–1915. https://doi.org/10.1016/j.stemcr.2017.10.018.

57. Chen, J., Rattner, A., and Nathans, J. (2005). The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes. Journal of Neuroscience *25*, 118–129. https://doi.org/10.1523/JNEUROSCI.3571-04.2005.

58. Baluapuri, A., Wolf, E., and Eilers, M. (2020). Target gene-independent functions of MYC oncoproteins. Nat Rev Mol Cell Biol *21*, 255–267. https://doi.org/10.1038/s41580-020-0215-2.

59. Liu, H., Zhang, Y., Zhang, Y.-Y., Li, Y.-P., Hua, Z.-Q., Zhang, C.-J., Wu, K.-C., Yu, F., Zhang, Y., Su, J., et al. (2020). Human embryonic stem cell-derived organoid retinoblastoma reveals a cancerous origin. Proceedings of the National Academy of Sciences *117*, 33628–33638. https://doi.org/10.1073/pnas.2011780117.

60. Pettker, C.M., Goldberg, J.D., El-Sayed, Y.Y., Copel, J.A., and ; Committee on Obstetric Practice, the American Institute of Ultrasound in Medicine, and the Society for Maternal-Fetal Medicine, D. (2017). Committee Opinion No 700: Methods for Estimating the Due Date. Obstetrics & Gynecology *129*.

61. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J *17*, 10. https://doi.org/10.14806/ej.17.1.200.

62. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

63. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol *33*, 290–295. https://doi.org/10.1038/nbt.3122.

64. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res *4*, 1521. https://doi.org/10.12688/f1000research.7563.1.

65. Orjuela, S., Huang, R., Hembach, K.M., Robinson, M.D., and Soneson, C. (2019). ARMOR: An A utomated R eproducible MO dular Workflow for Preprocessing and Differential Analysis of R NA-seq Data. G3 Genes|Genomes|Genetics *9*, 2089–2096. https://doi.org/10.1534/g3.119.400185.

66. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol *36*, 411–420. https://doi.org/10.1038/nbt.4096.

67. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

68. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv. https://doi.org/10.48550/ARXIV.1802.03426.

69. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol *37*, 38–44. https://doi.org/10.1038/nbt.4314.

70. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment *2008*, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

71. Alasoo, K. (2019). wiggleplotr: Make read coverage plots from BigWig files. R package version 1.8.0. Preprint.

72. Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res *22*, 2008–2017. https://doi.org/10.1101/gr.133744.111.

73. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol *38*, 276–278. https://doi.org/10.1038/s41587-020-0439-x.

74. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

75. Svoboda, M., Frost, H.R., and Bosco, G. (2022). Internal oligo(dT) priming introduces systematic bias in bulk and single-cell RNA sequencing count data. NAR Genom Bioinform *4*, lqac035. https://doi.org/10.1093/nargab/lqac035.

76. Ibrahim, M.M., and Kramann, R. (2019). genesorteR: Feature Ranking in Clustered Single Cell Data. bioRxiv, 676379. https://doi.org/10.1101/676379.

77. Blighe, K., Rana, S., and Lewis, M. (2018). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. Preprint.

78. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res *47*, W199–W205. https://doi.org/10.1093/nar/gkz401.

79. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

80. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

81. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci Rep *7*, 16878. https://doi.org/10.1038/s41598-017-17204-5.

82. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell Detection with Star-Convex Polygons. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag), pp. 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.

83. Cheng, H. (2004). Photoreceptor-specific nuclear receptor NR2E3 functions as a transcriptional activator in rod photoreceptors. Hum Mol Genet *13*, 1563–1575. https://doi.org/10.1093/hmg/ddh173.

84. Lerner, L.E., Gribanova, Y.E., Ji, M., Knox, B.E., and Farber, D.B. (2001). Nrl and Sp Nuclear Proteins Mediate Transcription of Rod-specific cGMP-phosphodiesterase β-Subunit Gene. Journal of Biological Chemistry *276*, 34999–35007. https://doi.org/10.1074/jbc.M103301200.

85. Lee, S., and Cobrinik, D. (2020). Improved third-generation lentiviral packaging with pLKO.1C vectors. Biotechniques *68*, 349–352. https://doi.org/10.2144/btn-2019-0155.

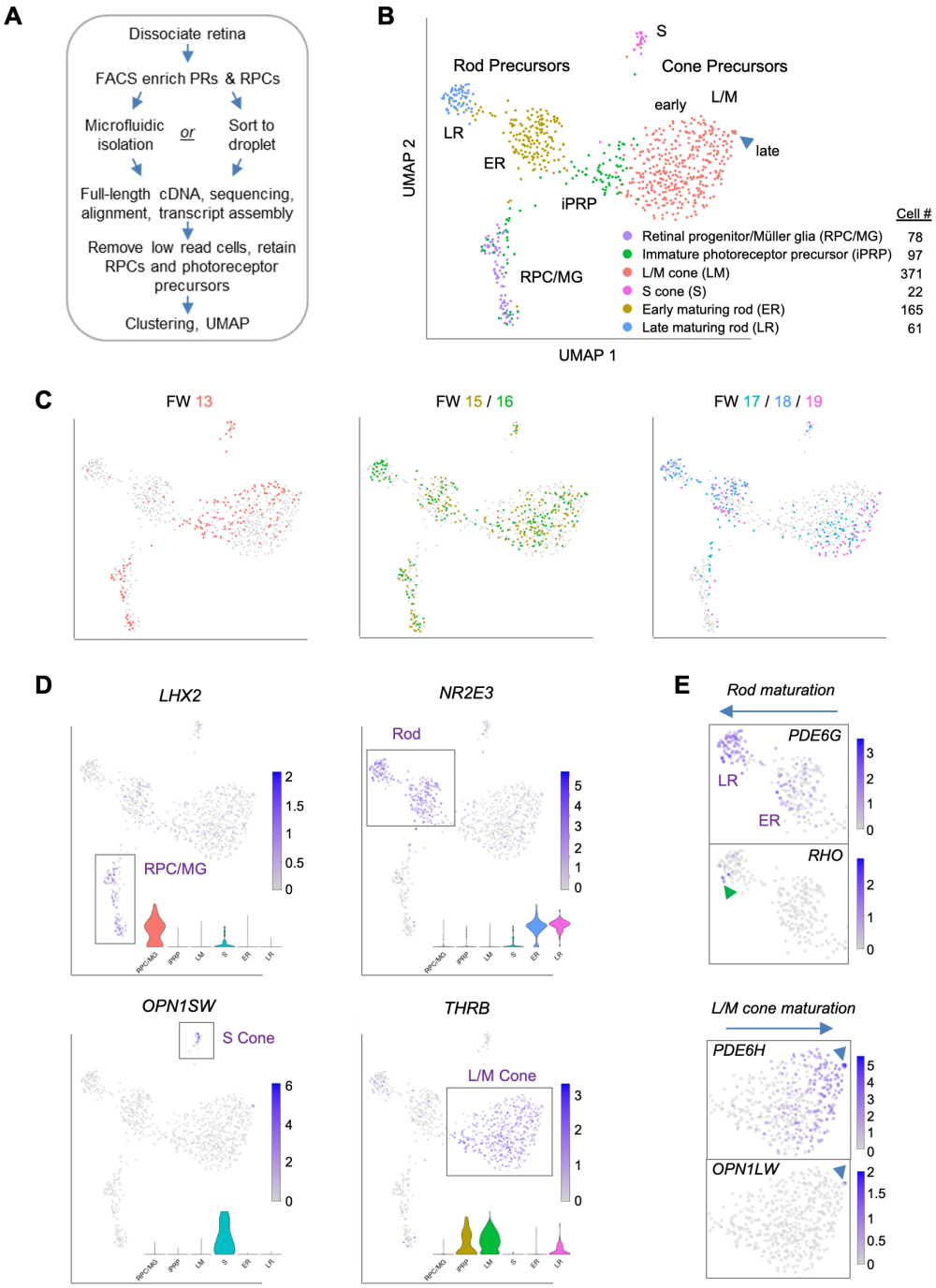


Figure 1: Photoreceptor-enriched full-length scRNA-seq of developing human retina. A. Overview of sample collection and sequencing. **B, C.** UMAP plots showing low-resolution cell type clusters (B) and ages (C). **D.** Expression of marker genes for RPC/MGs (*LHX2)*, rods (*NR2E3)*, S cones (*OPN1SW)*, L/M cones (*THRB*). *Insets*: Gene expression violin plots (from *left* to *right*): RPC/MG (red), iPRP (brown), LM cone (green), S cone (teal), early rod (blue), late rod (pink). **E.** Expression of markers of rod maturation (*PDE6G*, *RHO*) and cone maturation (*PDE6H*, *OPN1LW*). Arrowheads: Late-maturing *RHO*+ rods (*top*), late-maturing *OPN1LW*+ cones (*bottom*). See Figure S2 for additional examples. UMAP and violin plots for any gene or transcript isoform can be produced at https://docker.saban.chla.usc.edu/cobrinik/app/seuratApp/.

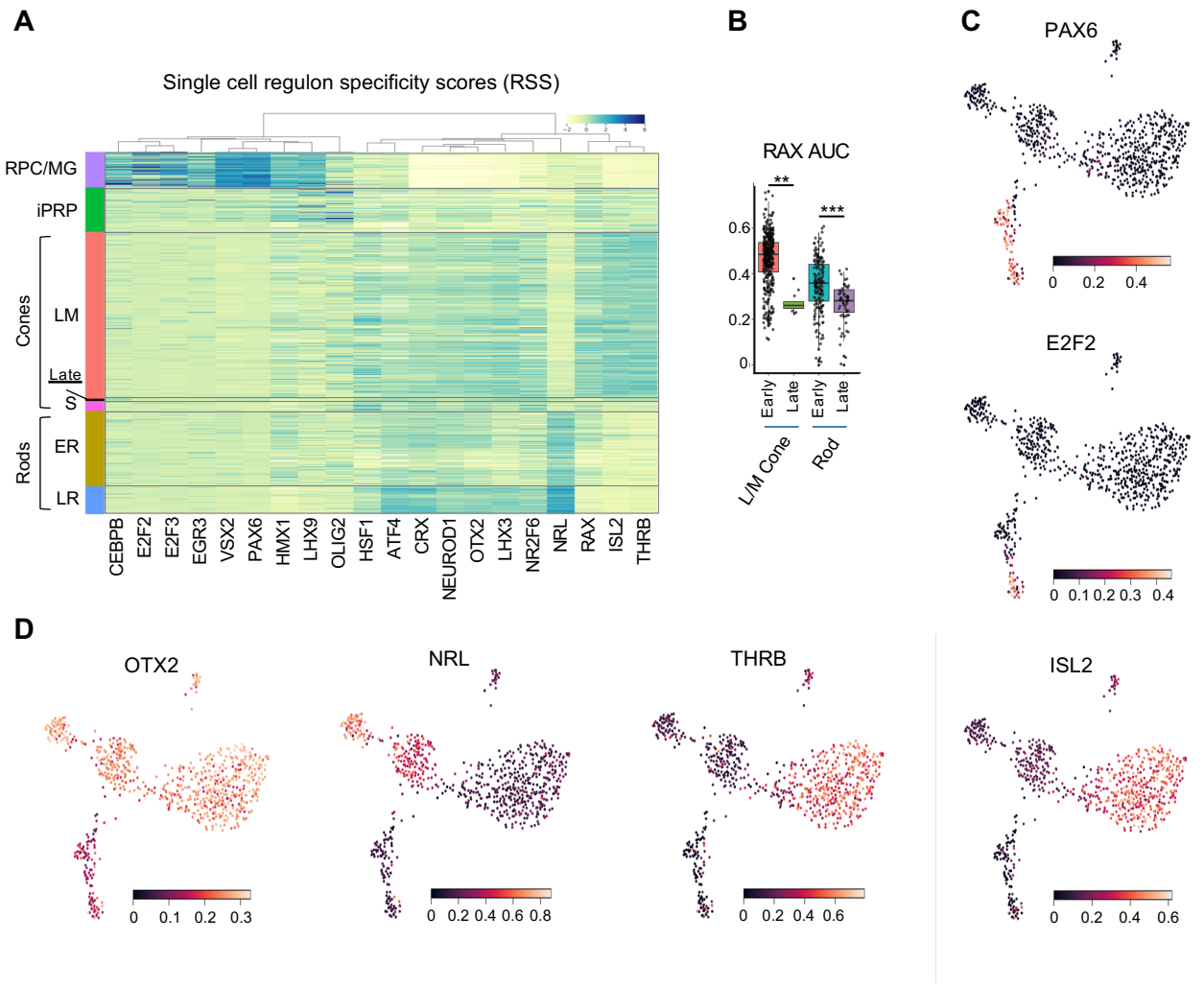


Figure 2: Regulon-defined RPC and photoreceptor precursor states A. Ward-clustered heatmap of the highest scoring SCENIC regulons in each cluster, displaying Z-score normalized regulon activities. Late = late-maturing L/M cones. **B.** Box plot of RAX regulon area under the curve (AUC) values for early and late L/M cones and rods. *, $p<0.005$; ***, $p<0.0005$, Dunn test. **C,D.** UMAP plots of regulon AUC values for (C) PAX6 (RPC/MG) and E2F2 (RPC), and (D) OTX2 (photoreceptors and photoreceptor-committed RPCs), NRL (rod) and THRB and ISL2 (L/M cone).

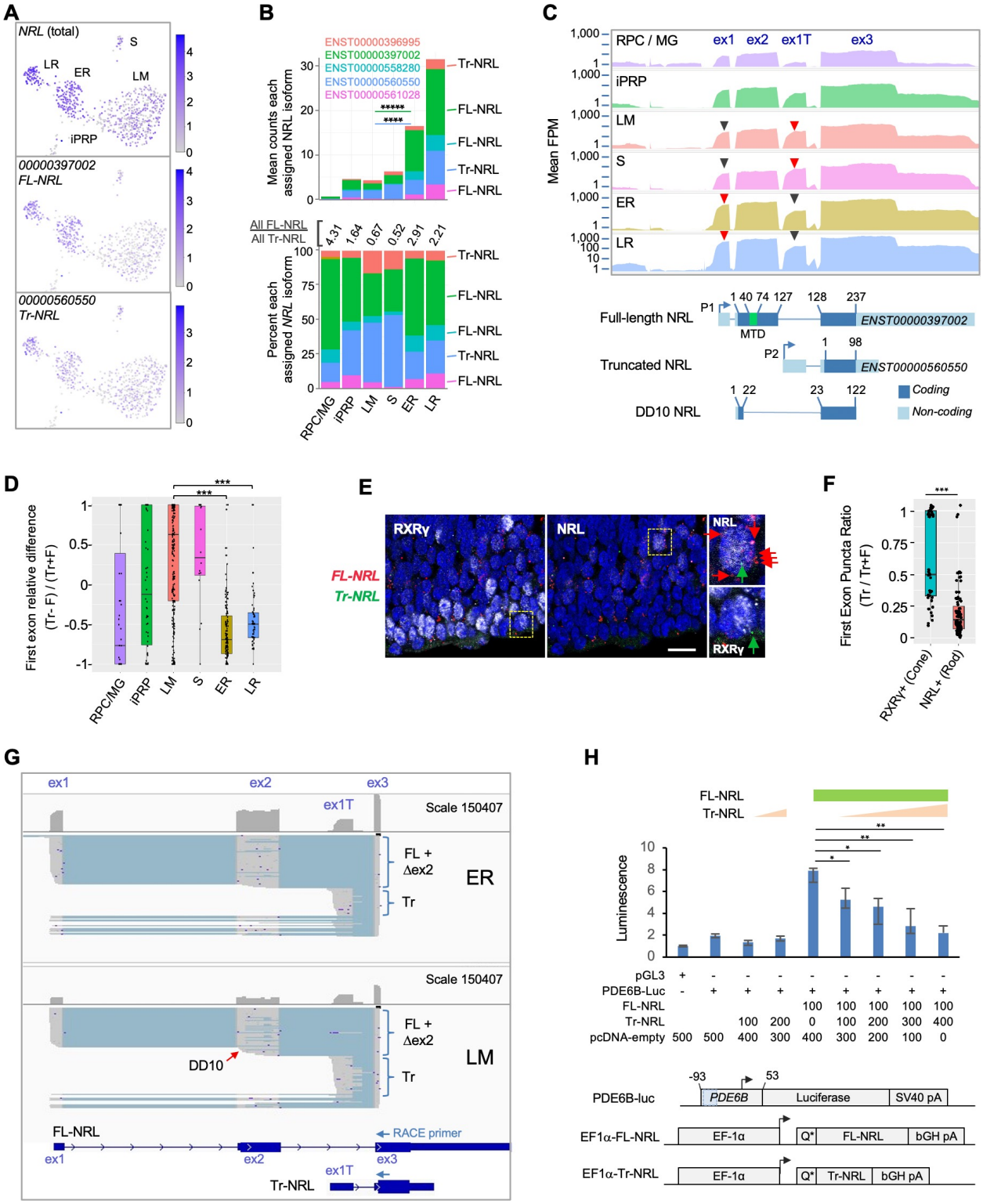


Figure 3: Differential expression of *NRL* isoforms in rod and cone precursors. A. Expression of *NRL* gene and the most highly assigned Ensembl isoforms *ENST00000397002* (FL-NRL) and *ENST00000560550* (Tr-NRL). **B.** Mean *NRL* isoform assignments for clusters defined in Figure 1B, presented as total counts (*top*) and percentage of total counts (*bottom*). Significance for LM

vs. ER fold change, colored by isoform. ****, p <0.0002; *****, <0.000001 (bootstrapped Welch's t-test). Ensemble transcript IDs shown in color with structures shown in Figure S4B. **C.** *Top*: Mean read counts (fragments per million, FPM) across Ensembl *NRL* exons for each cluster. *Bottom*: Transcript structures numbered according to amino acid positions. Minimal transactivation domain (MTD) in green. Arrowheads: Red/black: First exons where red is higher of two peaks. **D.** Relative difference box plot of raw reads mapping to truncated (Tr) and full length (F) transcript first exons in each cell, according to cluster. Relative difference is the difference in reads mapping to truncated and full-length *NRL* first exons (Tr-F) divided by the sum of both (Tr+F). Values >0 indicate more reads assigned to truncated isoform, values < 0 indicate more reads assigned to full-length isoform. ***, p <0.0001 (post-hoc Dunn test). **E.** NRL and RXRγ immunostaining and RNA FISH with probes specific to truncated Tr-NRL exon 1T (green puncta) and FL-NRL exons 1 and 2 (red puncta) in FW16 retina. Boxed regions enlarged at right show an RXRγ^{lo}, NRL^{+} rod with one Tr-NRL and six FL-NRL puncta (*top*) and an RXRγ^{hi}, NRL^{-} cone with one Tr-NRL and no FL-NRL puncta (*bottom*), indicated with same-color arrows. Scale bar: 10 µm. **F.** Ratio of fluorescent puncta observed in experiment depicted in (E) for NRL+ or RXRγ^{hi} cells where Tr puncta >0. ***; p <0.0005 (Welch's t-test). **G.** Long-read nanopore sequencing of pooled 5' RACE reactions initiated with *NRL* exon 3 primers and performed on cDNA libraries from 23 ER cells (*top*) and 21 LM cells (*bottom*). Each schematic shows total exon coverage (*above*) and individual transcripts (*below)*, where expressed sequences are grey and introns light blue. Full-length (FL), alternatively spliced or internally initiated exon 2 (Δex2), and truncated (Tr) transcripts are indicated by brackets. Red arrow: Transcripts resembling DD10, with internal exon 2 transcription initiation and premature splicing to exon 3. Ensembl FL-NRL and Tr-NRL transcript isoforms and RACE primer positions are shown below. **H.** *Top*: PDE6B-luciferase reporter activity in NIH-3T3 cells transfected with indicated amounts (ng) of pcDNA4-C-EF1α and derived FL-NRL and Tr-NRL constructs. *Bottom:* PDE6B-luc reporter and pcDNA4-C-EF1α expression constructs. Blue box = NRL response element. Error bars = standard deviation of triplicate measurements. *, p <0.05; **, <0.005 (Student's t-test). Data representative of two experiments in NIH-3T3 and one in 293T.

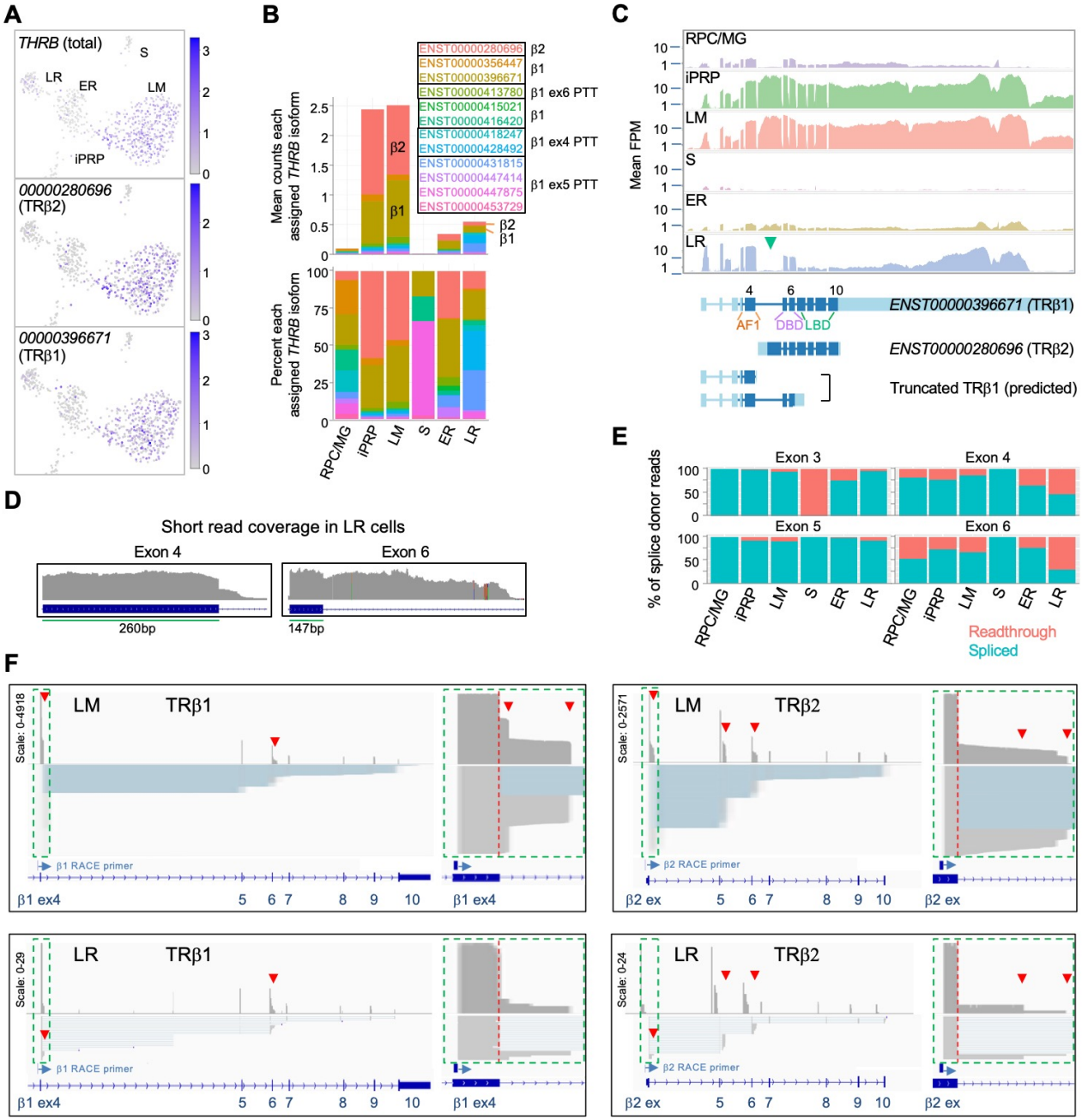


Figure 4: Differential expression of *THRB* isoforms in rod and cone precursors. A. Expression of *THRB* and highly assigned isoforms *ENST00000280696* (encoding TRβ2) and *ENST00000396671* (TRβ1). **B.** Mean *THRB* isoform assignments for each cluster presented as counts (*top*) and percentage of counts (*bottom*); Ensemble transcript IDs shown in color with β2, β1, and β1 PTT isoform structures as in Figure S6B. **C.** *Top*: Mean read counts across Ensembl *THRB* exons. *Bottom*: Transcript structures for TRβ1, TRβ2, and two TRβ1 truncations. Green arrowhead: First TRβ2 exon. *ENST00000396671* exon numbers are indicated above and protein domains (AF1, DNA-binding (DBD), and ligand binding (LBD)) below. **D.** Read coverage for LR cells across *THRB* exons 4 and 6 splice donor sites. **E.** Percentage of exon splice donor reads that are spliced or readthrough to the subsequent intron. **F.** Long-read sequencing of pooled 3' RACE reactions initiated with exon 4 (*left*) or TRβ2 exon 1 (right) performed on cDNA libraries

from 21 LM cells (*top*) and 5 LR cells (*bottom*). Schematics show total coverage (*above*) and individual transcripts (*below)*. TRβ1 and TRβ2 first exons (green boxes) are enlarged at right. Red arrowheads: intronic premature transcription termination (PTT).

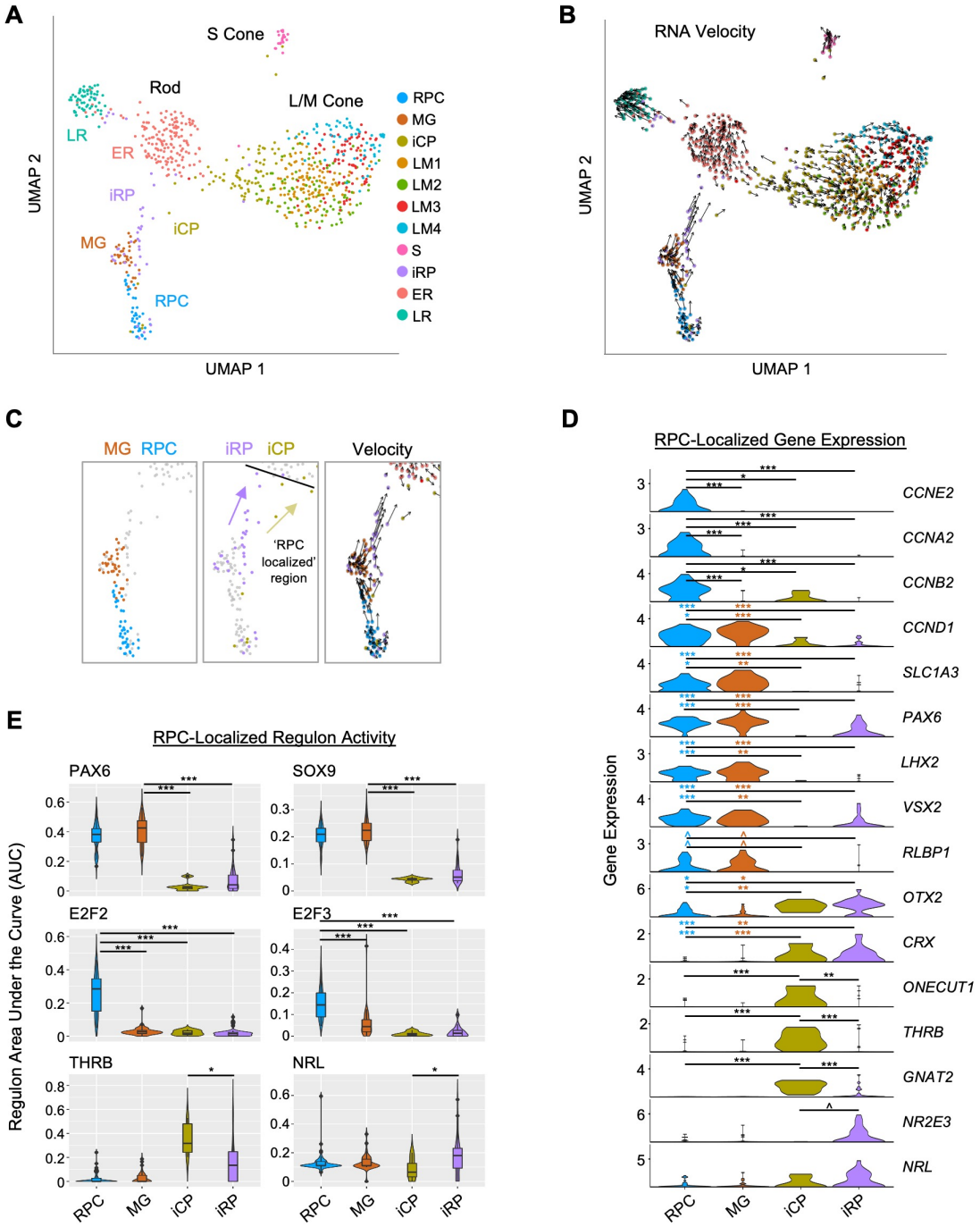


Figure 5: Two post-mitotic photoreceptor precursor populations expressing rod or cone markers. A. UMAP plot colored by high resolution clusters. **B.** RNA velocity plots with cell clusters as in A. **C.** Enlarged view highlighting RPC and MG clusters (*left*), RPC-localized iCP and iRP clusters (*middle*), and RNA velocity (*right*). Black line: limit of RPC-localized region. Arrows depict inferred trajectories. **D.** Violin plots depict expression of selected genes in RPC, MG, and RPC-localized iCP and iRP cells. Colored asterisks compare cluster of same color to cluster at right of line. **E.** SCENIC regulon violin and box plots for RPC-localized cells in each cluster, selected from most specific regulons for MG, RPC, ER and LM clusters. ^, pAdj <0.1; *, <0.05; **, <0.005; ***, <0.0005 (post-hoc Dunn test).

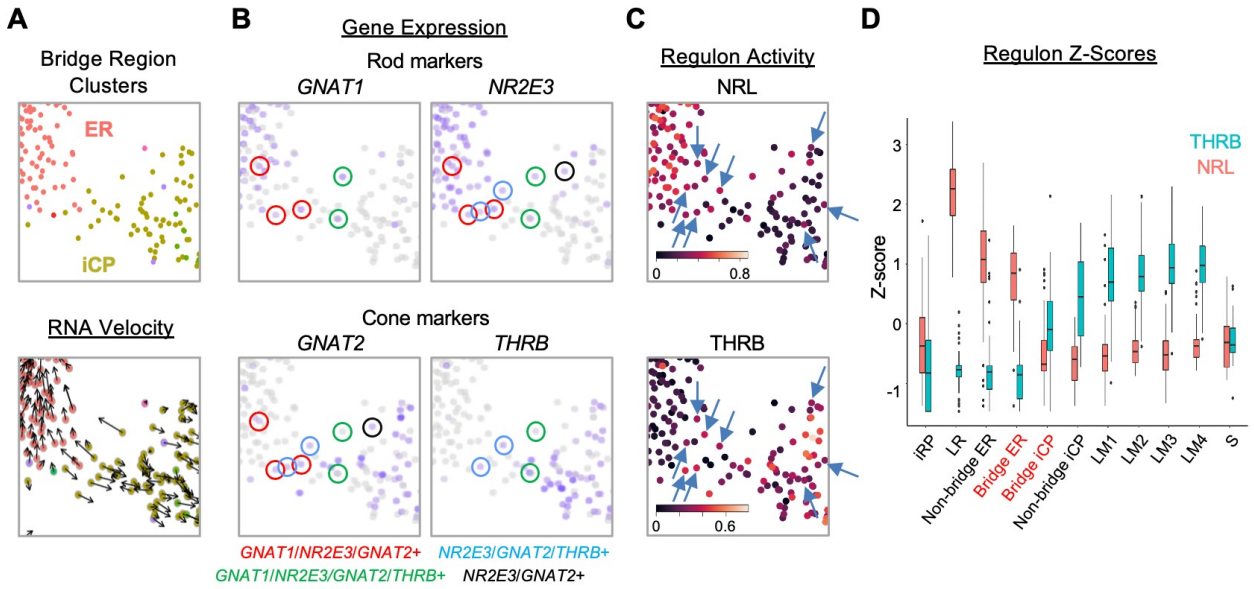


Figure 6: **An iCP sub-population with cone- and rod-related RNA co-expression. A-C.** UMAP 'bridge' region cells colored by ER and iCP cluster and RNA velocity (A), rod and cone marker gene expression (B), and NRL and THRB regulon activity (C, arrows indicate cells with both regulon signals). **D**. Box plot of Z-score-normalized NRL and THRB regulon AUCs for each cluster; Bridge ER and Bridge iCP represent cells present in the UMAP region in panels A-C.

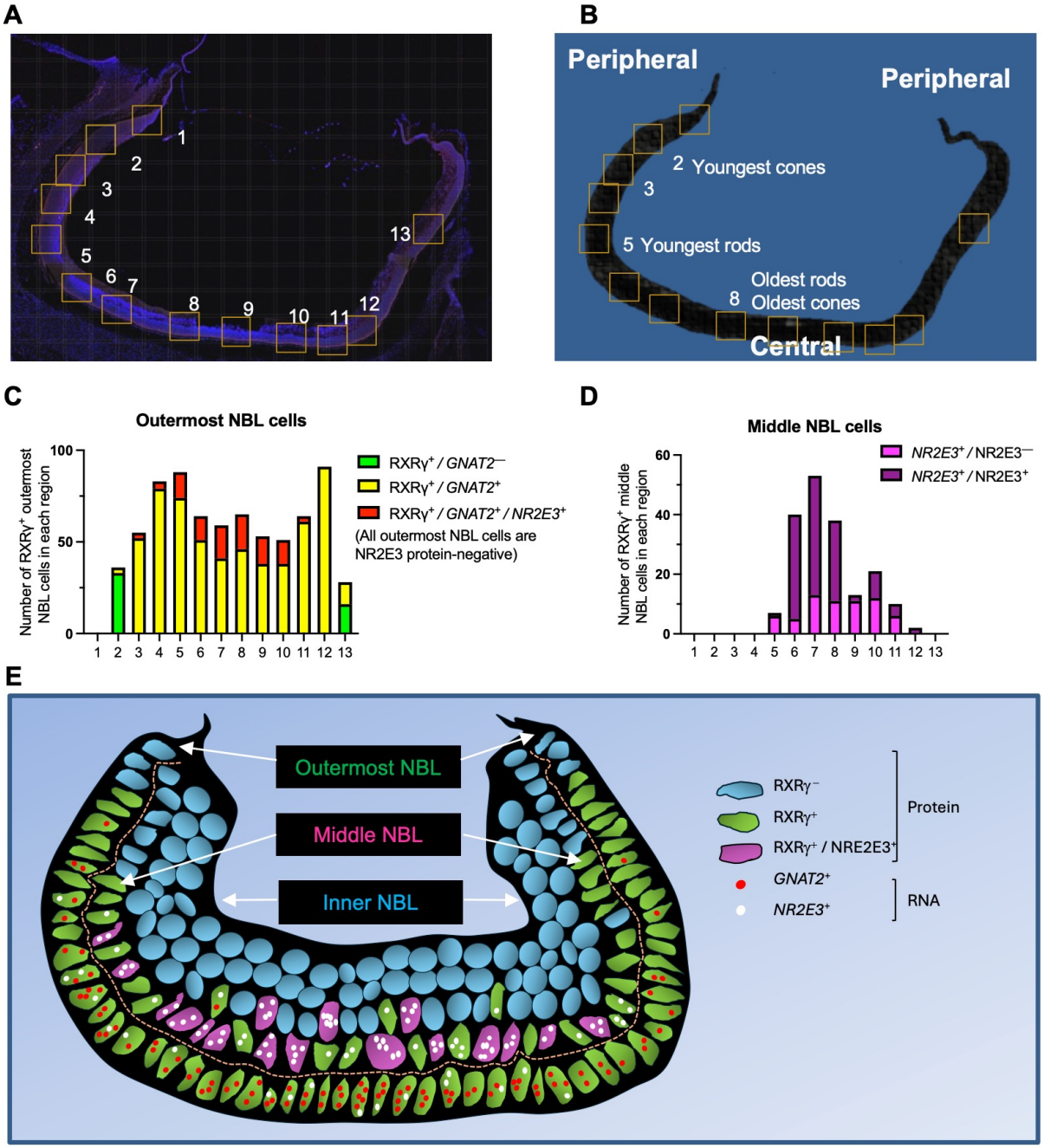


Figure 7: Cone-related *GNAT2* and rod-related *NR2E3* RNA co-expression in human cone precursors. A,B. Tiled composite fluorescence image of FW14 retinal section after immunofluorescence staining of RXRγ and NR2E3 and RNA FISH of *GNAT2* and *NR2E3* (A) and diagram of the same section indicating the most peripheral (*i.e.*, youngest and least mature) and most central (*i.e.*, oldest and most mature) rods and cones (B). Boxes indicate regions further evaluated in Supplementary Figure S12 and quantitated in panels C and D. **C,D.** Quantitation of (C) outermost and (D) middle (*i.e.*, sub-outermost) NBL photoreceptor precursors expressing combinations of RXRγ and NR2E3 proteins and *GNAT2* and *NR2E3* RNAs (*n.b.*, italics are used for RNAs, non-italics for proteins). **E.** Patterns of *GNAT2* and *NR2E3* RNA and RXRγ and NR2E3

protein expression inferred from *in situ* hybridization and immunofluorescence staining. RXRγ is expressed in the outermost NBL starting in the far periphery, consistent with cone precursors, and in middle NBL cells, starting more centrally, consistent with rod precursors. *GNAT2* and *NR2E3* RNA co-expression in outermost NBL cells lacked NR2E3 protein, representing putative cone precursors. RXRγ+ retinal ganglion cells in the inner NBL and ganglion cell layer not shown.

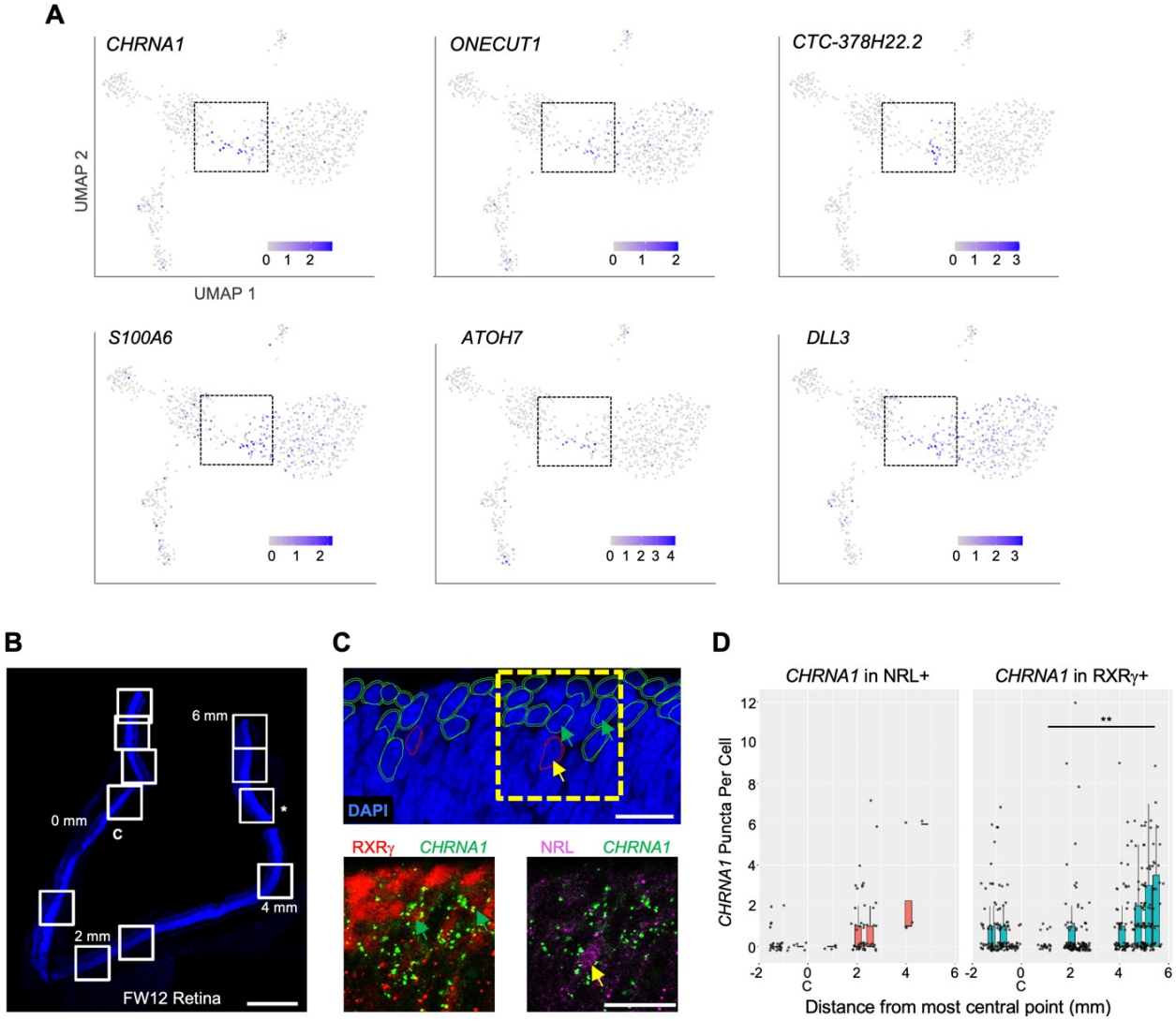


Figure 8. Developmental expression of early cone and rod precursor markers. A. UMAP plots of iCP marker genes *CHRNA1*, *ONECUT1*, *CTC-378H22.2*, and *S100A* (see Figure S9A) and previously identified photoreceptor precursor markers *ATOH7* and *DLL3*. **B-D.** Combined RXRγ/NRL immunohistochemical staining and *CHRNA1* RNA FISH of FW12 retina. **B**. Tiled images of retina section with nuclei stained with DAPI. White boxes: fields used for quantitative fluorescent imaging. Distances along apical edge of tissue marked in mm from midpoint of central image (0 mm, C). *: Imaged region shown in C. Scale bar = 500 µm. **C.** *Top:* Retinal nuclear and cellular segmentation and identification of cells as RXRγ+ (green outline) or NRL+ (red outline). Yellow box: Field shown below. *Bottom:* RXRγ or NRL immunofluorescence staining with *CHRNA1* FISH. Arrows: RXRγ+,*CHRNA1*+ (green), NRL+,*CHRNA1*+ (yellow). Scale bars = 15 µm. **D**. Quantitation of fluorescent puncta in RXRγ+ and NRL+ cells by image field. X-axis: Distance from the midpoint of each image to retina center (0 mm, C). **, p <0.005 (Wald test, images from 0-6 mm).

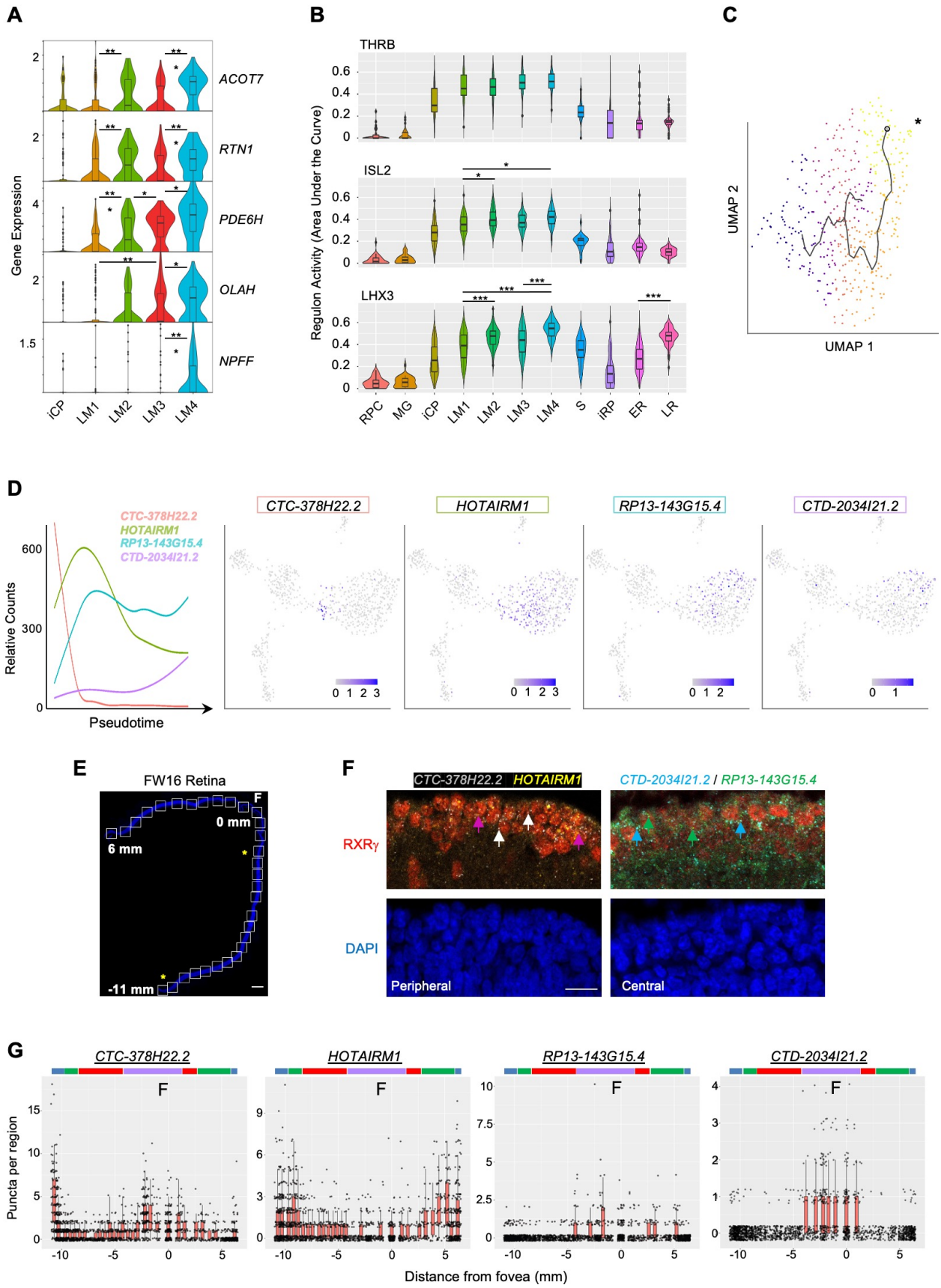

A
Gene Expression
ACOT7
RTN1
PDE6H
OLAH
NPFF
iCP
LM1
LM2
LM3
LM4
B
Regulon Activity (Area Under the Curve)
THRB
ISL2
LHX3
RPC
MG
iCP
LM1
LM2
LM3
LM4
S
iRP
ER
LR
C
UMAP 2
UMAP 1
D
CTC-378H22.2
HOTAIRM1
RP13-143G15.4
CTD-2034I21.2
Relative Counts
Pseudotime
E
FW16 Retina
F
0 mm
6 mm
-11 mm
F
CTC-378H22.2 HOTAIRM1
CTD-2034I21.2 / RP13-143G15.4
RXRγ
DAPI
Peripheral
Central
G
CTC-378H22.2
HOTAIRM1
RP13-143G15.4
CTD-2034I21.2
Puncta per region
Distance from fovea (mm)

Figure 9: L/M cone subcluster marker genes, regulons, and pseudotemporal trajectory with successive lncRNA gene expression. A. Violin plots of high-resolution cone cluster marker genes with increasing maturation-associated expression. All significant differences between adjacent clusters are indicated. **B.** Violin plots of regulons with highest LM1-4 cone cluster specificity scores. *, $p < 0.05$; **, <0.005; ***, <0.0005 (post-hoc Dunn test). **C.** Pseudotemporal trajectory through the L/M cone population derived with Monocle 3. *: Root cell used to define endpoint. The pseudotime trajectory may be related to LM1-LM4 subcluster distributions in Figure 5A. **D.** Trendlines of relative count expression (*left*) and UMAP plots for lncRNAs correlating with early or late-upregulating modules. Line color matched to labels. **E, F.** Combined RXRγ immunohistochemical staining and FISH of lncRNAs on FW16 retina. **E**. Tiled images of retina with nuclei stained with DAPI. White boxes: fields used for quantitative fluorescent imaging. Distances along apical edge of tissue marked in mm from fovea to ciliary margins. Scale bar = 500 µm. Asterisks identify fields shown in F. **F.** Combined RXRγ immunostaining and multiplex FISH for four lncRNAs, of which two are shown in peripheral and central retina regions. Arrows: White: RXRγ/*CTC-378H22.2*+. Magenta: RXRγ/*HOTAIRM1*+. Blue: RXRγ/*CTD-2034I21.2*+. Green: RXRγ/*RP13-143G15.4*+. Scale bar = 15 µm. **G.** Quantitation of lncRNA fluorescent puncta assigned to RXRγ+ cells after segmentation. Colored bars mark lncRNA expression regions as described in the text. Arrows indicate significant count differences between images ($p < 0.05$, Wald test).

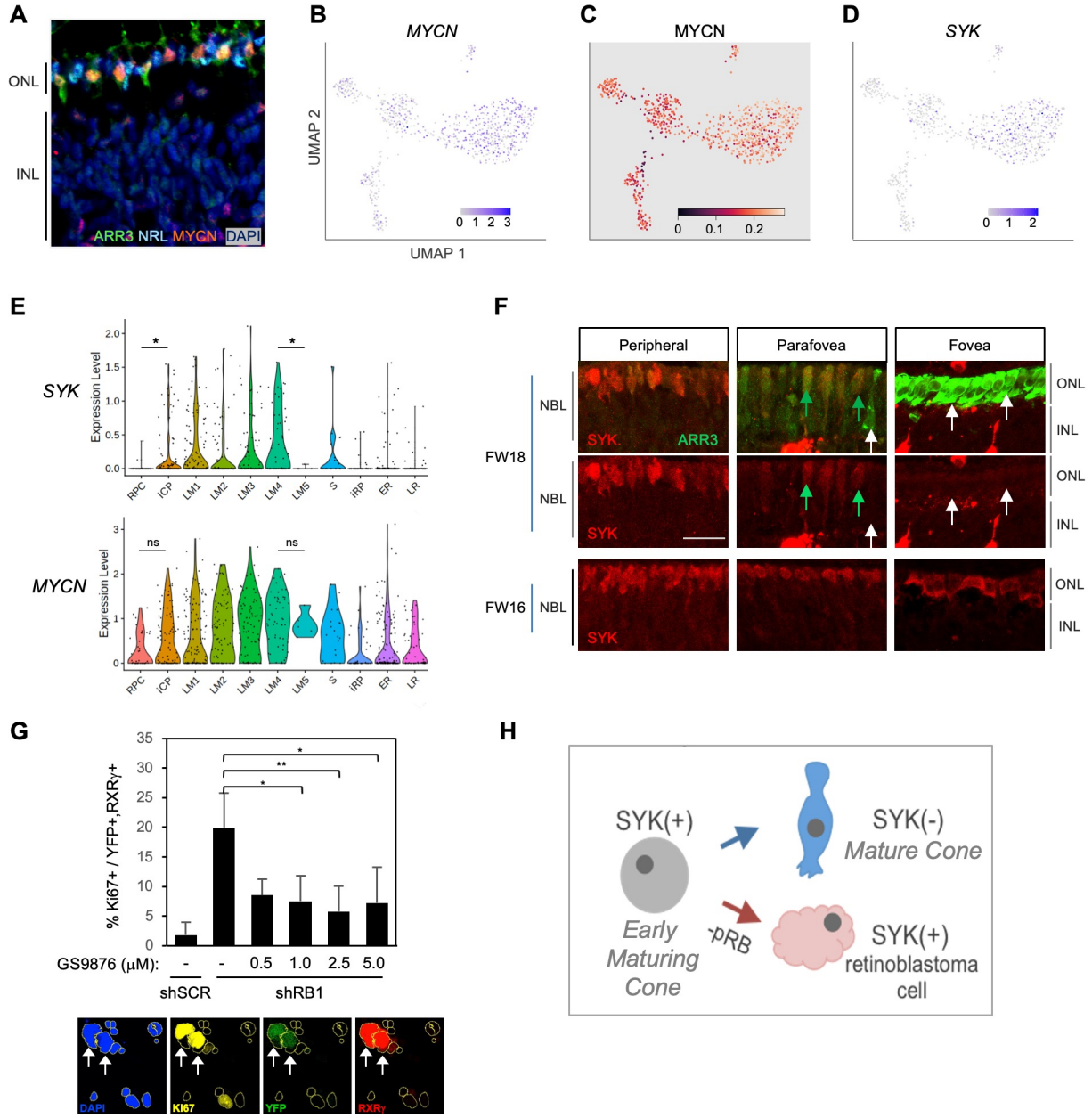


Figure 10: Cone intrinsic MYCN and SYK associated with proliferative response to pRB loss. A. Immunofluorescent staining shows high MYCN in ARR3+ cones but not in NRL+ rods in FW18 retina. **B-D.** UMAP plots of *MYCN* expression (B), MYCN regulon activity (C), and *SYK* expression (D). **E.** *SYK* and *MYCN* gene expression violin plots by cluster. *, p <0.05; ns = not significant (t-test). **F.** Immunohistochemical staining of SYK and cone arrestin (ARR3) in FW18 and FW16 retinae. Green arrow: ARR3+, SYK+. White arrow: ARR3+, SYK–. Scale bar = 25 μm. **G.** *Top:* Effect of SYK inhibitor GS-9876 on Ki67 expression in RXRγ+ cones from FW16.5 retina co-transduced with YFP and shRB1- or control shSCR-shRNA. Values represent means of three analyses from two treatment replicates. Error bars: standard deviation. *, p <0.05; **, <0.005 (Student's T-test with equal variance, 2-tailed). RXRγ+ cells: Experiment 1, n=1,340. Experiment 2, n=804. Range 107 – 366 cells per condition. *Bottom:* Example of Ki67, YFP, and RXRγ co-

immunostaining with DAPI+ nuclei (yellow outlines). Arrows: Ki67+, YFP+, RXRγ+ nuclei. **H.** Model of SYK expression in cone maturation and retinoblastoma development.

Supplementary Materials for

Identification and characterization of early human photoreceptor states and retinoblastoma-related features

Dominic W. H. Shayler et al.

Supplementary Figures:

Figure S1: scRNA-seq sample and sequencing summary.

Figure S2: Expression of marker genes of RPCs, Müller glia, photoreceptors, rods, and cones.

Figure S3: Differential expression between rod and cone maturation states.

Figure S4: Differential *NRL* transcript isoform expression in early rod and cone precursors.

Figure S5: Cell type-specific NRL protein expression.

Figure S6: Differential *THRB* isoform assignments in rod and cone precursors.

Figure S7: THRB isoforms and premature transcription termination in cone and rod precursors.

Figure S8: *RXRG* isoform expression in rod and cone precursors.

Figure S9: High resolution cluster marker genes and differential RPC versus MG gene expression.

Figure S10: Immature photoreceptor precursor populations identified with deep full-length scRNA-seq and 3' snRNA-seq.

Figure S11: Cone and rod precursor cell type assignments, UMAP positions, and gene expression in a large 3' snRNA-seq analysis.

Figure S12: *NR2E3* and *GNAT2* co-expression in photoreceptor precursors.

Figure S13: Developmental expression of photoreceptor precursor markers in a large 3' snRNA-seq dataset.

Figure S14: iCP-enriched regulon activities.

Figure S15: Heatmaps of coregulated gene modules across cone precursor pseudotime.

Figure S16: Differential gene expression in early cone and rod precursors.

Figure S17: Proposed cell state trajectories in human L/M cone and rod development.

Key Resources Table:

Information related to strains, cell lines, biological samples, antibodies, recombinant DNA reagents, sequence-based reagents, commercial assays or kits, and software.

Supplementary Tables: Source data Excel files for selected figures.

Table S1: Differentially expressed genes for low resolution rod clusters.

Table S2: Differentially expressed genes for low resolution cone clusters.

Table S3: Differentially expressed genes for low resolution late- vs early-maturing L/M cone populations.

Table S4: Differentially expressed genes in rods vs rod precursors in a 3' snRNA-seq dataset (data of Zuo *et al*. PMID 39117640.

Table S5: Differentially expressed genes in cones vs cone precursors in a 3' snRNA-seq dataset (data of Zuo *et al*. PMID 39117640.

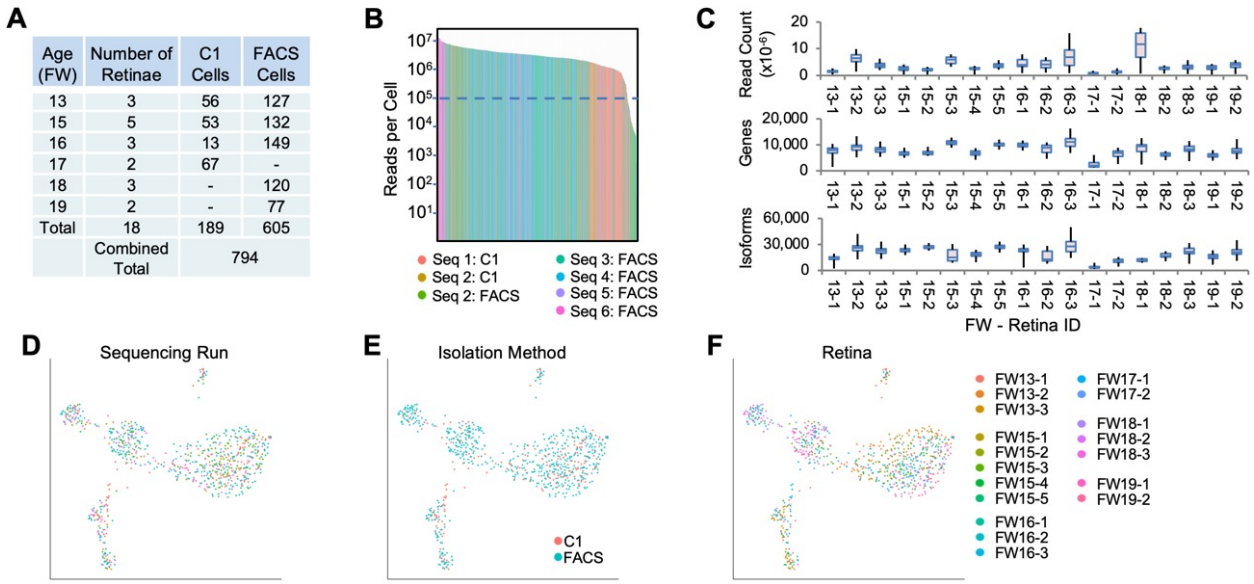
Table S6: SCENIC regulon specificity scores for low resolution clusters.

Table S7: Differentially expressed genes for high resolution MG and RPC clusters.

Table S8: SCENIC regulon specificity scores for high resolution clusters.

Table S9: Pseudotime-correlated genes and gene modules.

Table S10: Differentially expressed genes for low resolution early rod and L/M cone clusters.



Age (FW)	Number of Retinae	C1 Cells	FACS Cells
13	3	56	127
15	5	53	132
16	3	13	149
17	2	67	-
18	3	-	120
19	2	-	77
Total	18	189	605
	Combined Total	794	

Figure S1: scRNA-seq sample and sequencing summary. A. Retinae and cell numbers examined at each age. **B.** Histogram of total read counts for each cell, colored by sequencing run. Seq 1 cells were isolated by C1, Seq 2 cells by FACS or C1 as indicated, and all others by FACS. Dotted line: 100,000 read cutoff for cell exclusion. **C.** Box plots of read counts, genes detected, and Ensembl transcript isoforms detected per cell ordered by fetal age and specimen number. **D-F.** UMAP plots colored by sequencing run (D), isolation method (E), or retina ID (F).

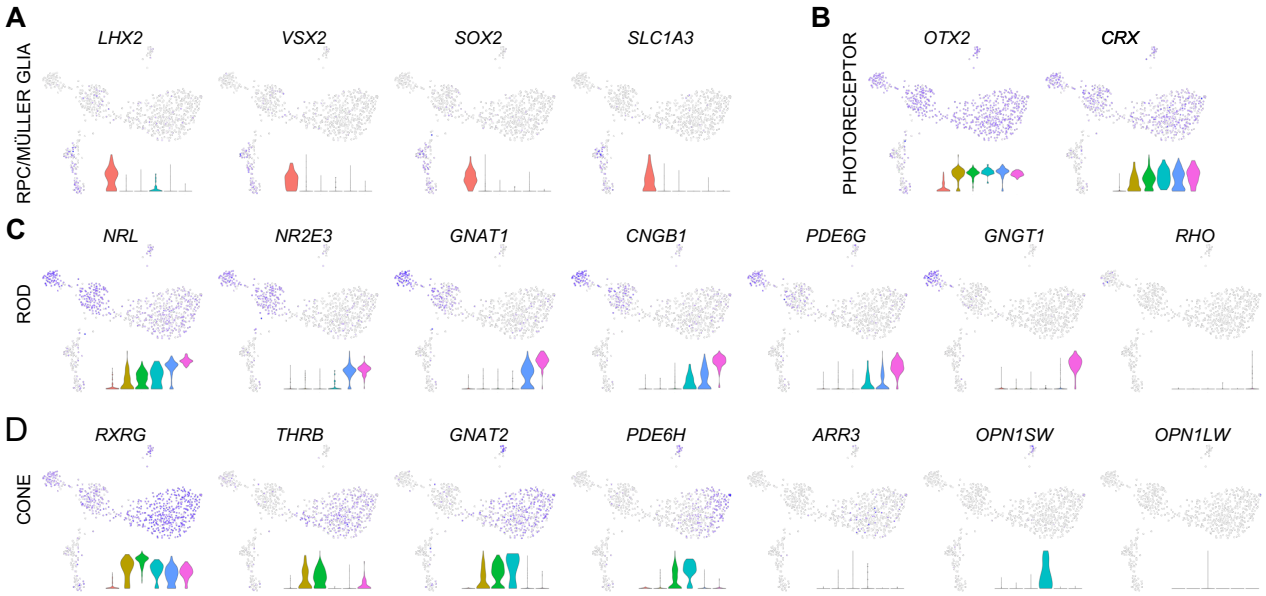


Figure S2: Expression of marker genes of RPCs and Müller glia (A), photoreceptors (B), rods (C), and cones (D). Insets: Gene expression violin plots (from *left* to *right*): RPC/MG (red), iPRP (brown), LM cone (green), S cone (teal), early rod (blue), late rod (pink). UMAP and violin plots for any gene or Ensembl transcript can be produced at https://docker.saban.chla.usc.edu/cobrinik/app/seuratApp/.

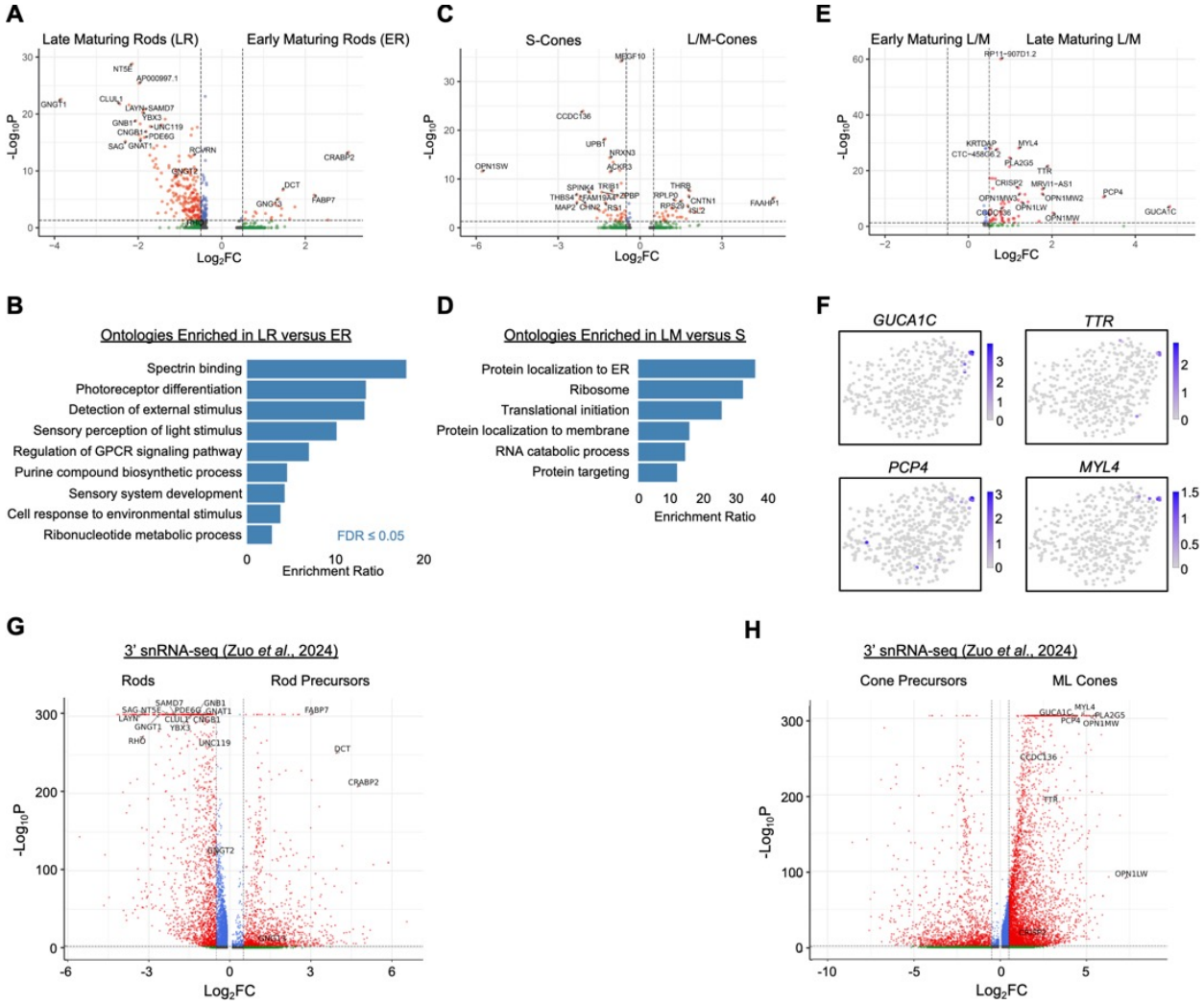


Figure S3. Differential expression between rod and cone maturation states. A, C, E. Volcano plots of differential expression (pAdj <0.05, $\log_2$FC >|0.5|) between ER and LR (A), LM and S (C), and spatially separated late-maturing L/M cones and remaining LM cluster (E). Labels indicate genes with highest significance and fold change. **B,D.** Overrepresentation of molecular function ontologies for genes upregulated in LR over ER (B) or upregulated in LM over S (D). **F.** UMAP plots of upregulated genes in the late-maturing L/M cone group. **G, H.** Volcano plots of differential expression (pAdj <0.01, $\log_2$FC >|0.5|) between Rod and Rod Precursor (G) and ML Cone and Cone Precursor (H) subclasses as defined Figure S3 of Zuo *et al*[20]. Late rod- and late cone-enriched genes labeled in (A) and (E) are also labeled when present in (G) and (H).

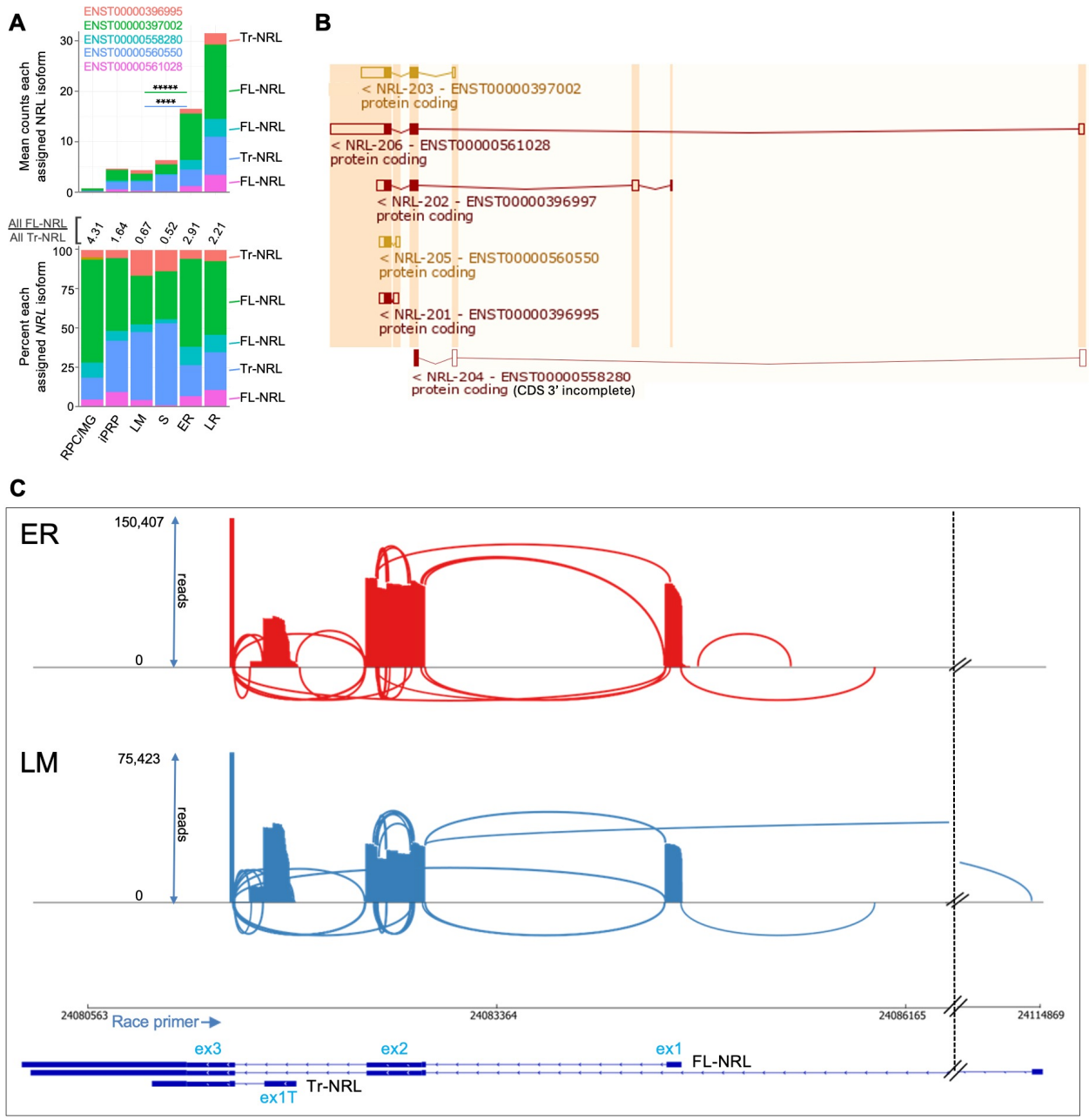


Figure S4: Differential *NRL* transcript isoform expression in early rod and cone precursors. **A.** Mean *NRL* isoform counts for each cluster based on short-read full-length cDNA sequencing (*top*) and percentage of total counts (*bottom*), as in Figure 3B. **B.** *NRL* Ensemble isoforms. **C.** *Top:* Sashimi plots of *NRL* transcript splicing based on nanopore long read sequencing of 5' RACE reactions from 23 ER cells and 21 LM cells. The minimum number of reads required to display transcript isoforms was set at 20 for ER cells and at 10 for LM cells, proportional to the total NRL reads in each sample. *Bottom*: Exon structures of FL-NRL (*ENST00000397002*) and Tr-NRL (*ENST00000560550*) isoforms and RACE primer location. The *NRL* gene is oriented relative to

chromosome 14 coordinates. Hash marks indicate sequences preceding a rarely used far upstream first exon in *ENST00000561028*. Note increased intra-exon 2 splicing and ex1T use in the LM population.

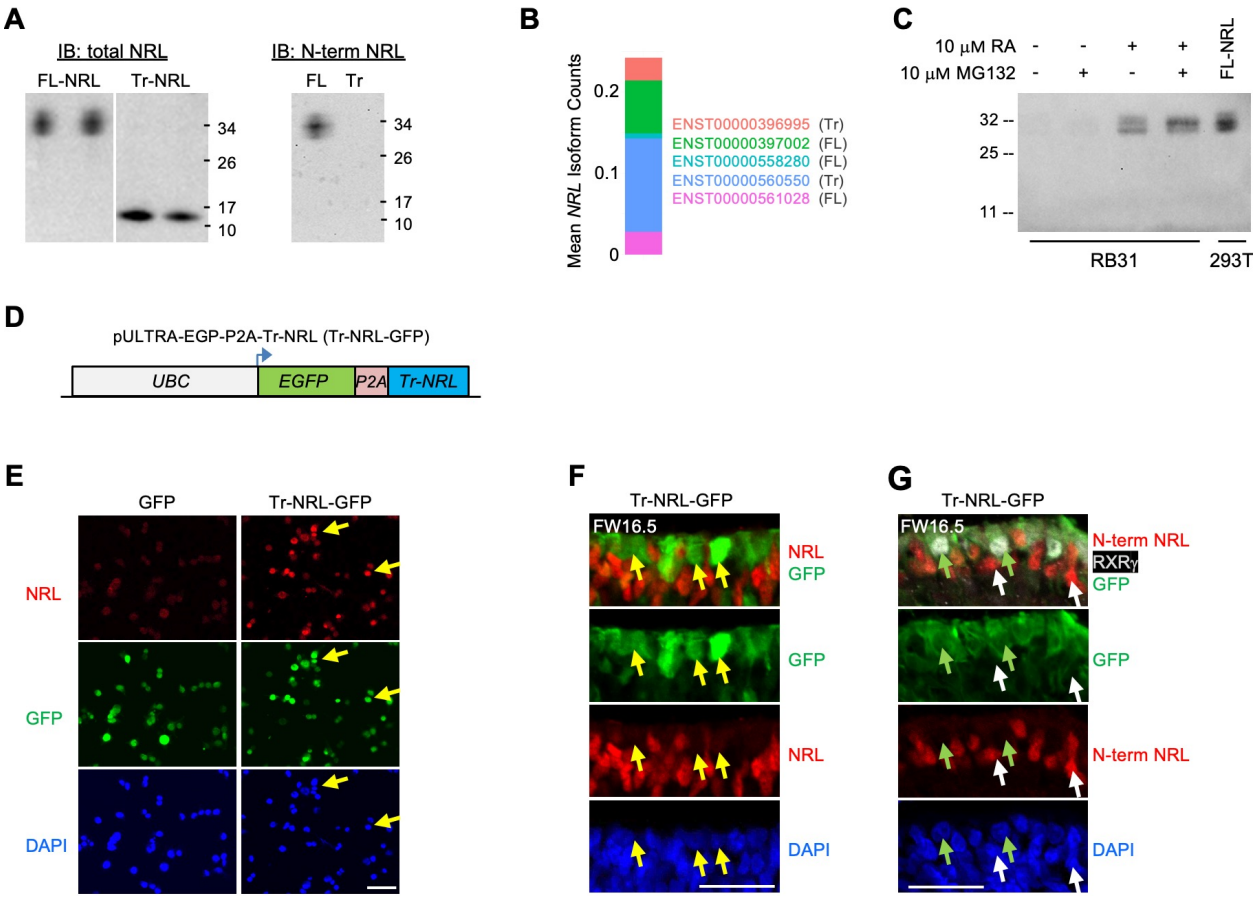


Figure S5: Cell type-specific NRL protein expression. A. NRL antibody validation by immunoblot (IB) of ectopic FL-NRL and Tr-NRL in HEK293T. *Left:* anti-total NRL. *Right:* anti-N-terminal NRL specific to FL-NRL. **B.** Average assigned *NRL* isoform expression in scRNA-seq of RB31 retinoblastoma cell line. **C.** Immunoblot of endogenous total NRL in RB31 cells treated with or without 10 μM retinoic acid (RA) and 10 μM proteasome inhibitor MG132 and in FL-NRL transfected HEK293T. **D.** pULTRA-EGFP-P2A-Tr-NRL lentiviral vector used for explanted retina transduction. **E.** NRL Immunofluorescent staining of HEK293T cells 48 h after lentiviral transduction of Tr-NRL-GFP. Scale bar = 50 μm. **F,G.** Immunostaining of total NRL (F) or N-terminal-NRL (G) in explanted FW16.5 fetal retina 7 d after lentiviral transduction of Tr-NRL-GFP. Arrows: Yellow = GFP+, NRL-. Green = GFP+,RXRγ+. White = GFP-,FL-NRL+. Scale bar, 25 μm.

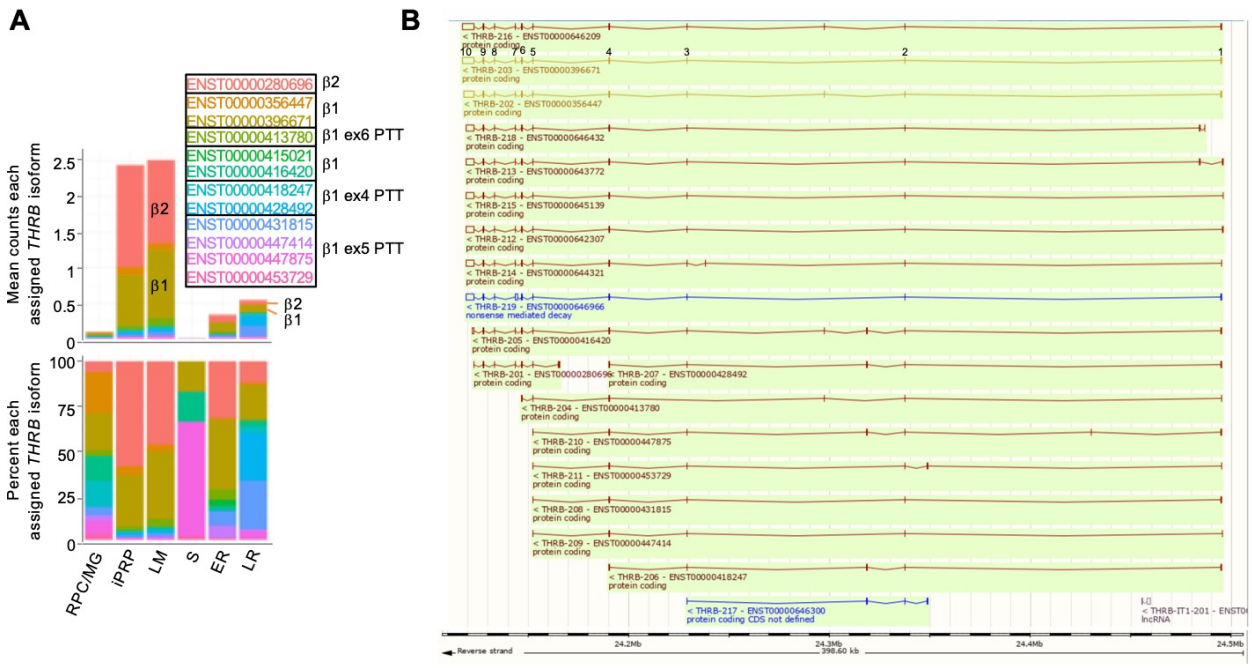


Figure S6: Differential *THRB* isoform assignments in rod and cone precursors. A. Mean *THRB* isoform assignments for each cluster presented as total transcript counts (*top*) and percentage of total counts (*bottom*) as in Figure 4B, with isoforms grouped according to capacity to encode full-length canonical TRβ1 or TRβ2 or to have PTT following ex4, ex5, or ex6 (exons numbered as for *ENST00000396671*) regardless of differences in 5' noncoding exons and 3' poly(A) sites. **B**. *THRB* Ensemble isoforms identified in panel A with *ENST00000396671* exon numbers indicated, accessed June 22, 2024, at: http://useast.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000151090;r=3:24117153-24495756.

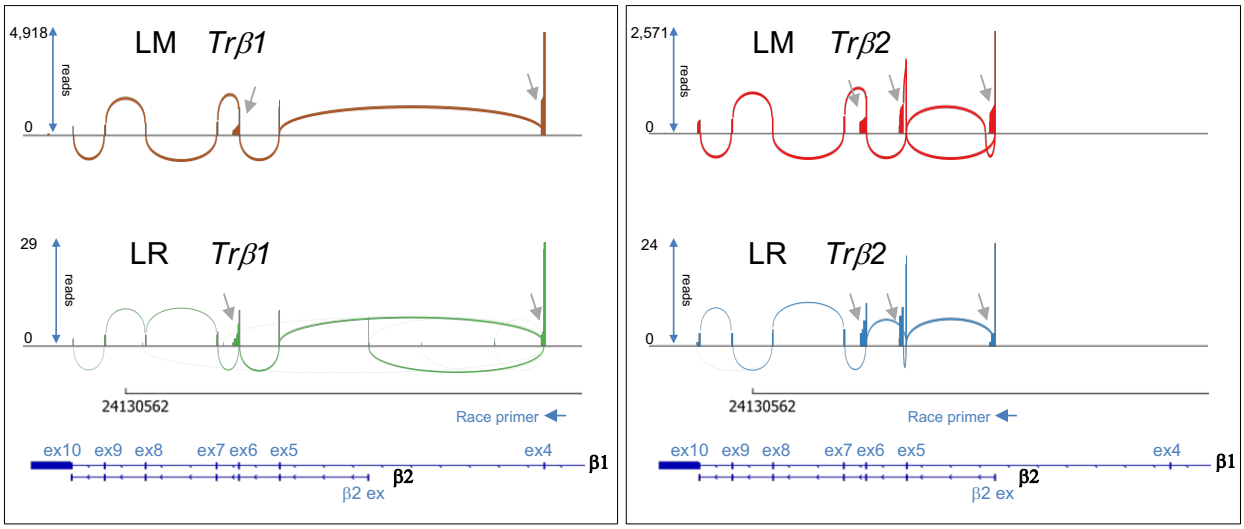


Figure S7: *THRB* isoforms and premature transcription termination in cone and rod precursors. *Top*: Sashimi plots of *THRB* transcript splicing based on long read nanopore sequencing of 3' RACE reactions from 23 LM cells and 5 LR cells. RACE reactions were initiated from primers specific to TRβ1 exon 4 (*left*) or specific to the unique TRβ2 exon (*right*). Gray arrows: premature transcription termination following TRβ1 exon 4, the TRβ2 first exon, the shared exon 6, and a novel exon used solely in TRβ2 transcripts. *Bottom*: TRβ1 and TRβ2 exon structures and RACE primer positions. The *THRB* gene is oriented relative to chromosome 3 coordinates.

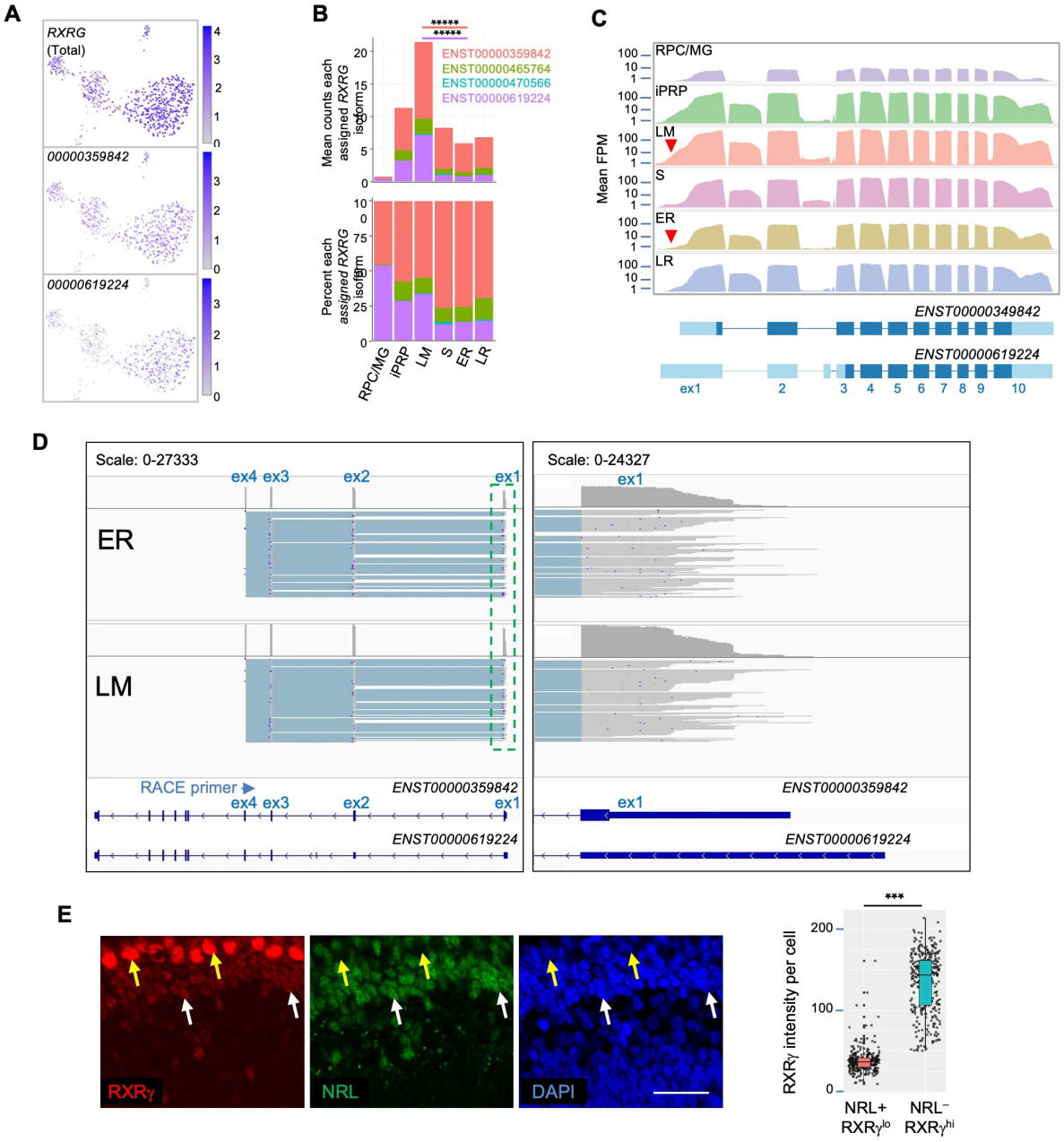


Figure S8: *RXRG* isoform expression in rod and cone precursors. A. Expression of *RXRG* and the two most highly assigned *RXRG* isoforms (*ENST00000359842* and *ENST00000619224*). **B.** Mean *RXRG* isoform assignments for each cluster presented as total counts (*top*) and as a percentage of total counts (*bottom*). Lines indicate significant LM vs. ER fold change, colored by isoform. *****, p <0.000001 (bootstrapped Welch's t-test). **C.** *Top*: Mean read counts across Ensembl *RXRG* exons. *Bottom*: Structures for *ENST00000359842* and *ENST00000619224.* Red arrowheads: Extended *RXRG* 5' UTR. **D.** Long-read nanopore sequencing of pooled 5' RACE initiated with *RXRG* exon 4 primers from 23 ER cells (*top*) and 21 LM cells (*bottom*). Schematic shows total exon coverage (*above*) and individual transcripts (*below)*. Panels at *left* show *RXRG*

locus with exon 1 in green box enlarged at right, revealing a similar range of RNA 5' ends in ER and LM populations. **E.** Immunohistochemical analysis of RXRγ and NRL in FW16 retina. Arrows: Yellow = RXRγ-high cone. White = NRL+ rod with weak RXRγ. Scale bar = 25 μm. *Right*: Boxplot of RXRγ mean grey values in NRL^{+},$RXR\gamma^{lo}$ and NRL^{-},$RXR\gamma^{hi}$ cells. ***, $p<0.0001$ (Welch's T-test).

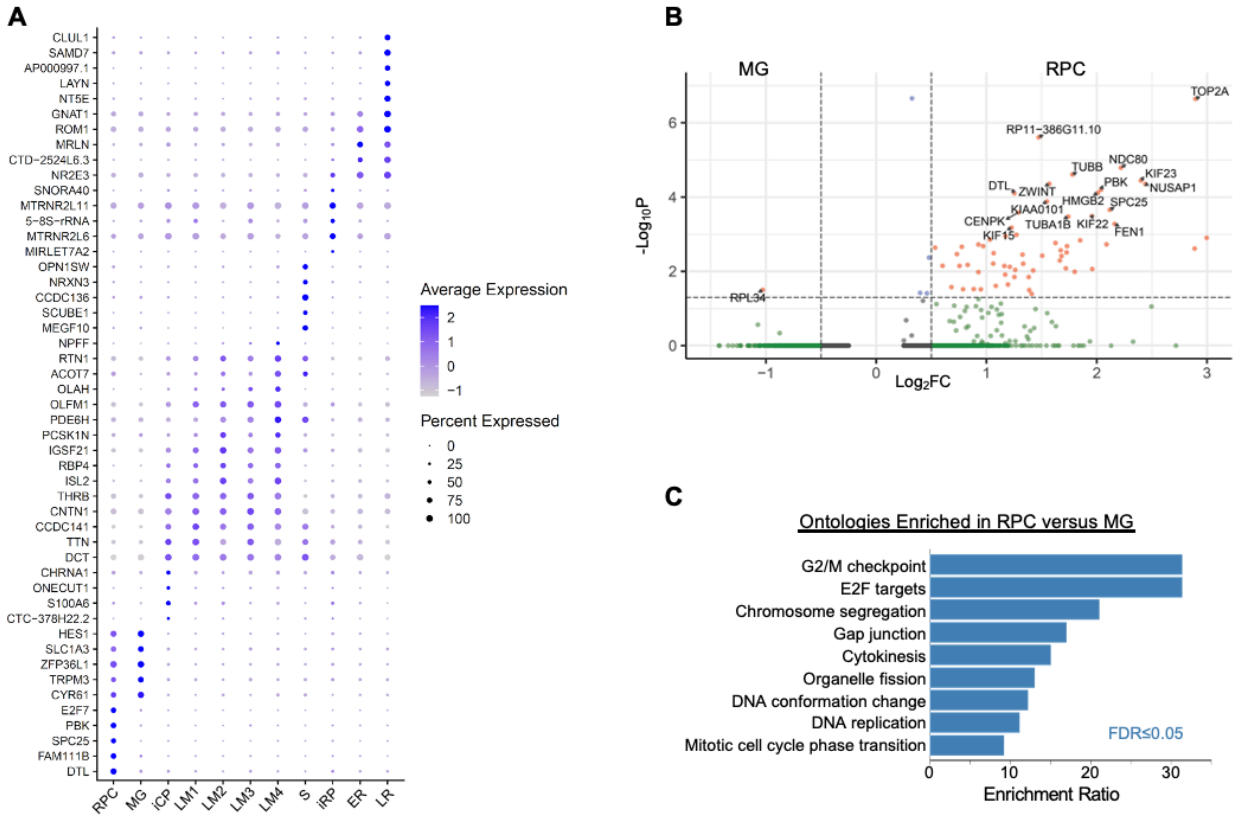


Figure S9: High resolution cluster marker genes and differential RPC versus MG gene expression. A. Marker gene dot plot for high resolution Louvain clusters. Dot size: Percent of cells expressing gene. **B.** Volcano plot of differential expression (pAdj <0.05, $\log_2$FC >|0.5|) between MG and RPC clusters with genes labeled when p $<10^{-4}$ except for *RPL34*. **C.** Overrepresentation analysis for genes upregulated in RPC above pAdj and $\log_2$FC cutoff.

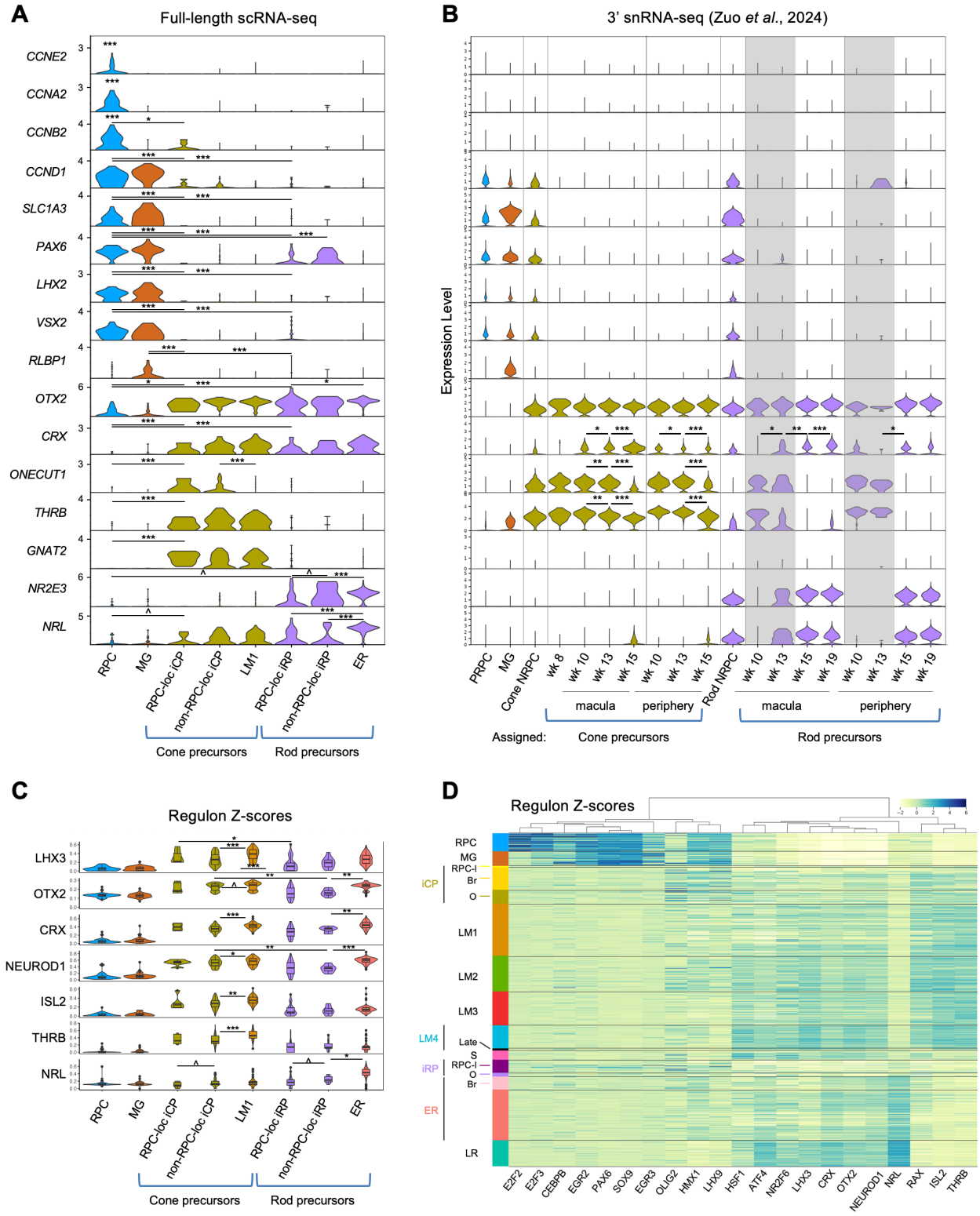
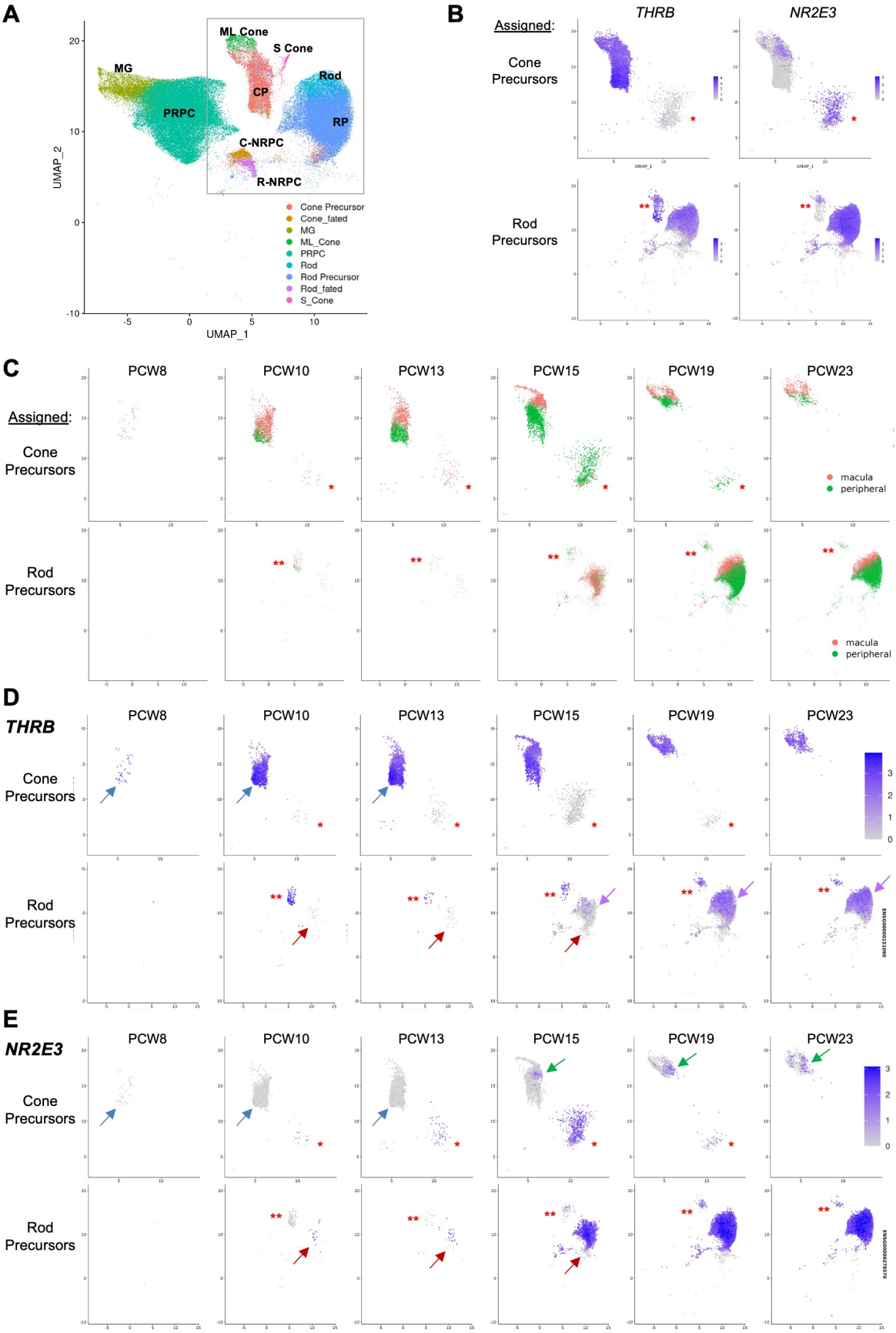


Figure S10: Immature photoreceptor precursor populations identified with deep full-length scRNA-seq and 3' snRNA-seq. A. Violin plots based on full-length scRNA-seq depicting gene expression in RPCs, MG, RPC-localized and non-RPC-localized iCPs and iRPs, and the early

maturing LM1 cone and ER clusters as designated in Figure 5A,C. **B.** Violin plots based on 3' snRNA-seq depicting expression of the same genes as in (A) in PRPCs, MG, cone- and rod-fated NRPCs, and cone and rod precursors at the indicated ages and macular versus peripheral retinal positions as designated in Zuo *et al,* 2024[20]. Grey boxes = potential mixed cone plus rod precursor populations as shown in Figure S11. **C,D**. Regulon activities based on full-length scRNA-seq in high-resolution RPC and photoreceptor precursor states. **C.** Violin plot depicting selected Z-score normalized regulon activities in the indicated cell populations. Significant differences ($p<0.05$) between corresponding iCP and iRP states are indicated for pan-photoreceptor regulons LHX3, OTX2, CRX, and NEUROD1, and significant differences between successive iCP – LM1 and iRP – ER states are indicated for all regulons. Activities of all regulons except NRL also increased in RPC-localized iCPs and iRPs relative to RPCs ($p<0.05$ for all). **D.** Ward-clustered heatmap of the highest scoring SCENIC regulons in each high-resolution cluster, displaying Z-score normalized regulon activities. Subcluster labels: RPC-l = RPC-localized, Br = Bridge-localized, O = Other/remainder of original cluster, Late = late-maturing LM4 cones. ^<0.1; *, <0.05; **, <0.005; ***, <0.0005 (post-hoc Dunn test).


A
ML Cone
S Cone
MG
Rod
CP
PRPC
RP
C-NRPC
R-NRPC
UMAP_2
UMAP_1
Cone Precursor
Cone_fated
MG
ML_Cone
PRPC
Rod
Rod Precursor
Rod_fated
S_Cone
B
Assigned:
THRB
NR2E3
Cone Precursors
Rod Precursors
C
PCW8
PCW10
PCW13
PCW15
PCW19
PCW23
Assigned:
Cone Precursors
Rod Precursors
macula
peripheral
D
THRB
Cone Precursors
Rod Precursors
E
NR2E3
Cone Precursors
Rod Precursors

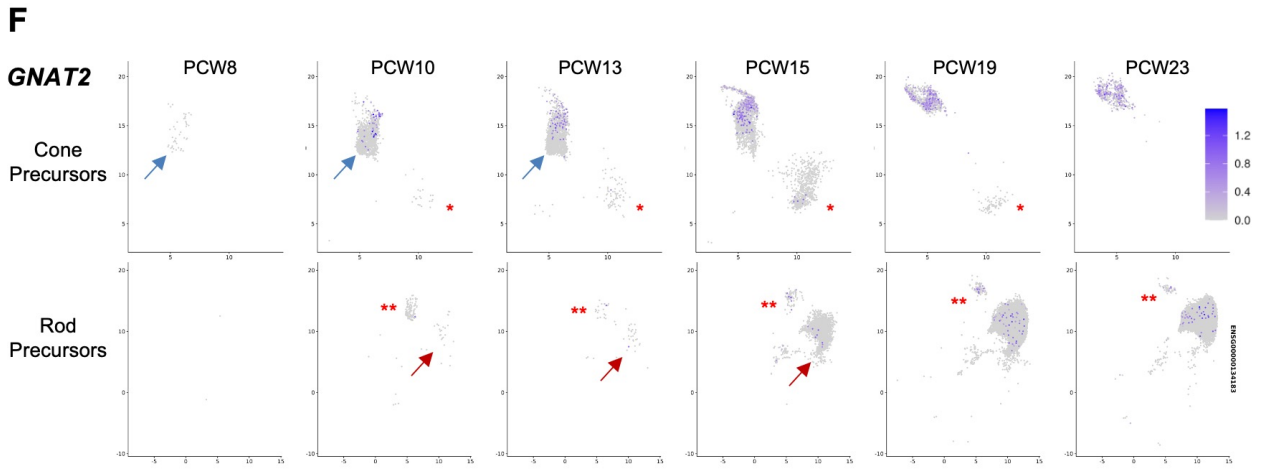


Figure S11: Cone and rod precursor cell type assignments, UMAP positions, and gene expression in a large 3' snRNA-seq analysis. A. UMAP positions of cell populations as assigned in Zuo et al, 2024[20]. MG = Muller glia; PRPC = primary retinal progenitor cells; C- and R-NRPC = cone and rod neurogenic retinal progenitor cells, CP = cone precursor, RP = rod precursor. Box = region enlarged in Cone Precursor UMAP plots in panels B-F. **B**. Expression of *THRB* and *NR2E3* overlaying UMAP plots of pcw 8-23 cone or rod precursors. ***** = possible rod precursors mis-assigned as cone precursors based on UMAP position, high *NR2E3*, and minimal *THRB*. ****** = possible cone precursors mis-assigned as rod precursors based on UMAP position, high *THRB*, and low *NR2E3*. **C**. UMAP plots of assigned cone and rod precursors in macula (orange) and peripheral retina (green) at each age. **D-F.** Expression of *THRB* (**D**), *NR2E3* (**E**), and *GNAT2* (**F**) overlaying UMAP plots of assigned cone or rod precursors at each age. Blue arrows = major cone precursor population at pcw 8, 10, 13 expressing *THRB* but not *NR2E3*. Red arrows = major rod precursor population at pcw 10, 13, 15 expressing *NR2E3* but not *THRB*. Purple arrows = later rod precursors with *THRB* co-expression at pcw 15, 19, 23. Green arrows = later cone precursors with *NR2E3* co-expression at pcw 15, 19, 23. Scale bars at right apply to all plots in each panel. Asterisks in **C-F** are as in panel **B**.

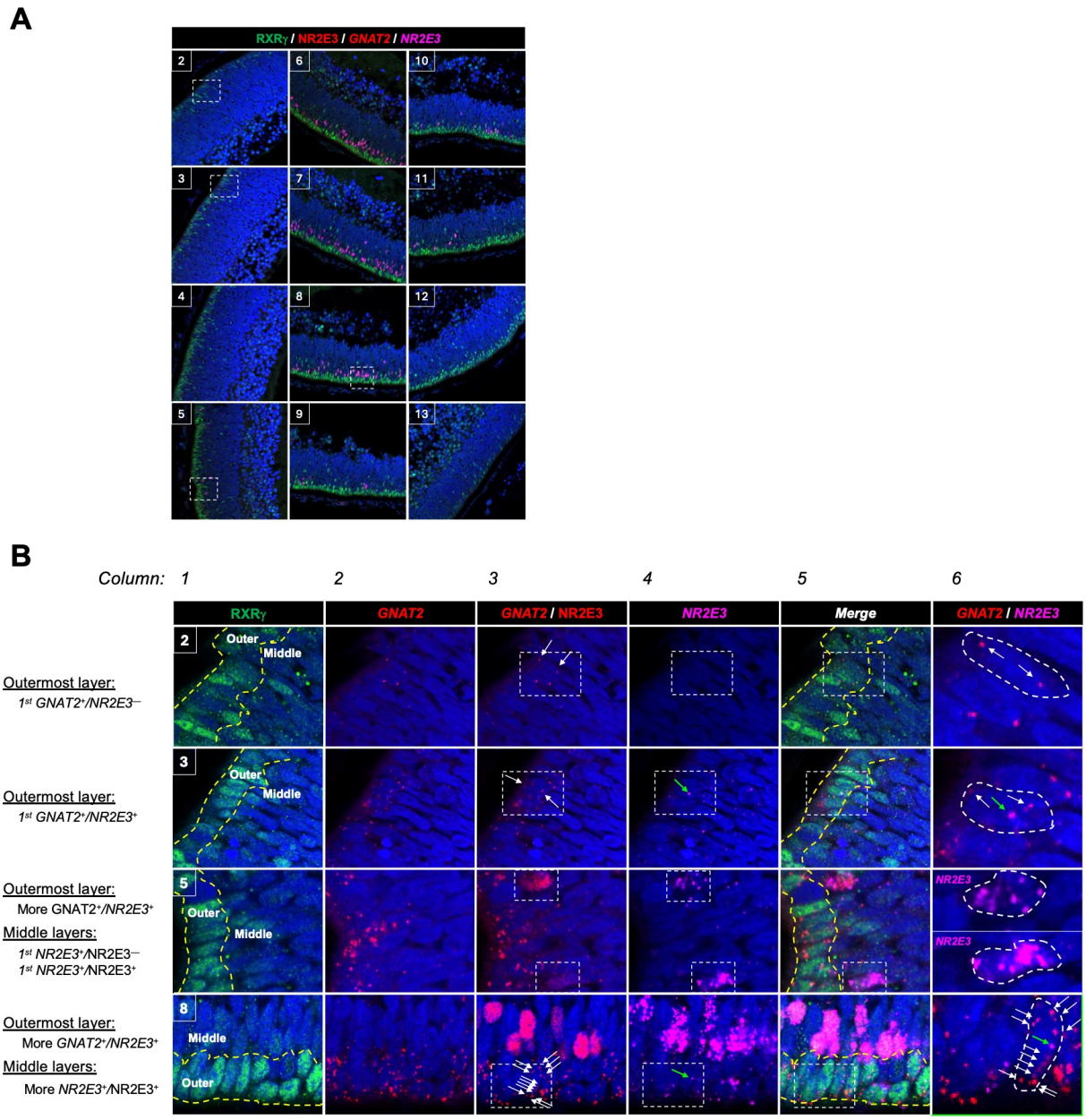


Figure S12: ***NR2E3* and *GNAT2* co-expression in photoreceptor precursors. A.** Composite images of combined RXRγ and NR2E3 immunofluorescence staining and *GNAT2* and *NR2E3* RNA FISH in boxed regions 2-13 of Figure 7A. **B.** Enlarged images of boxed regions in images 2, 3, 5, and 8 in panel A, as indicated at upper left of images in column 1. Dotted yellow lines demarcate the outermost and middle (sub-outermost) NBL cells analyzed separately in Figure 7C,D. White arrows = *GNAT2* puncta in boxed regions of interest. Green arrows = *NR2E3* puncta. Boxed regions in column 3 (*GNAT2* RNA plus NR2E3 protein) and column 4 (*NR2E3* RNA) are enlarged in column 6 (*GNAT2* RNA plus *NR2E3* RNA) and illustrate **region 2**) initial expression

of *GNAT2* RNA without *NR2E3* RNA in outermost NBL nascent cones; **region 3**) initial co-expression of *GNAT2* and *NR2E3* RNA in early outermost NBL cones; **region 5**) initial expression of *NR2E3* RNA with or without co-expressed NR2E3 protein; and **region 8**) co-expression of high *GNAT2* and low *NR2E3* RNA without NR2E3 protein in RXRγ^{hi} outermost NBL cones.

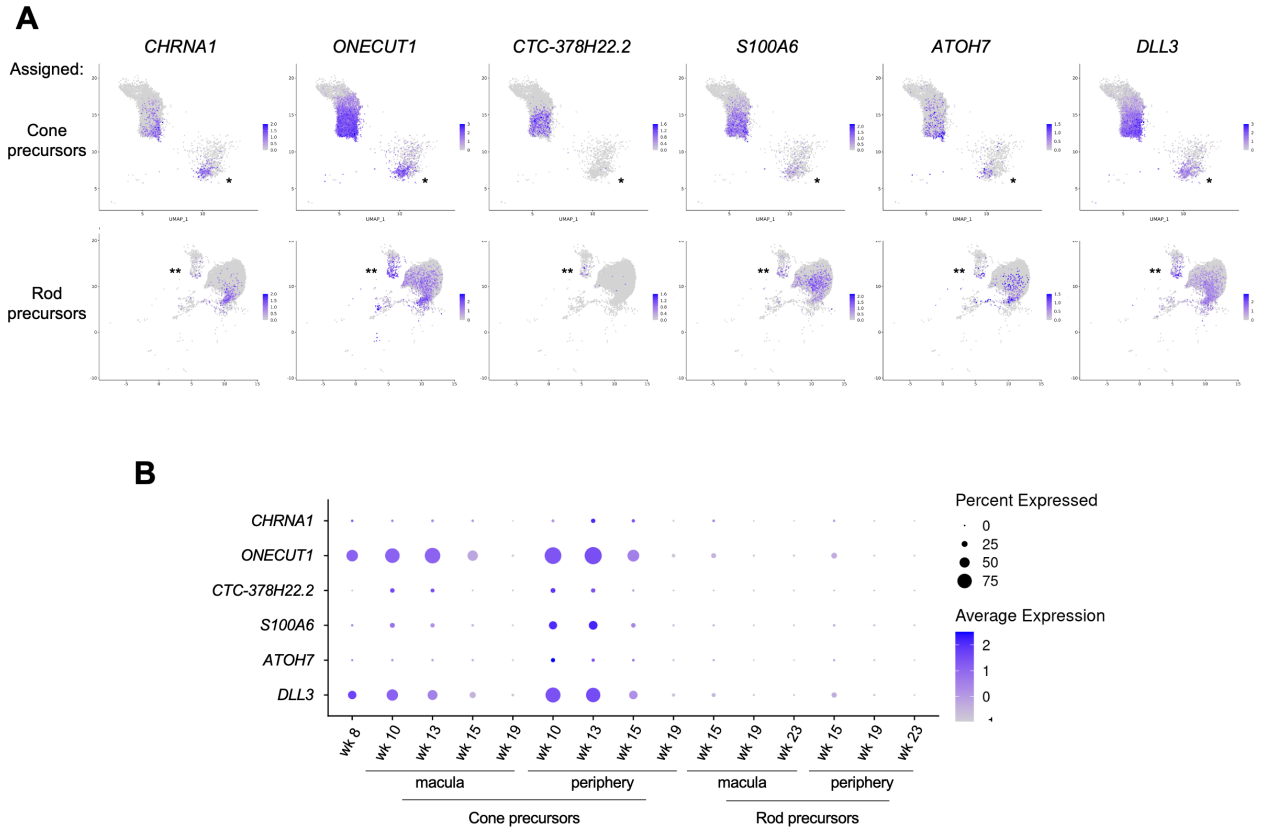


Figure S13: **Developmental expression of photoreceptor precursor markers in a large 3' snRNA-seq dataset. A.** Expression of selected markers in pcw 8-23 cone precursors (*top*) and rod precursors (*bottom*), as assigned in Zuo *et al*[20]. ***** = possible rod precursors mis-assigned as cone precursors; ****** = possible cone precursors mis-assigned as rod precursors. **B.** Dot plot depicting the percentage of cells expressing each gene and average expression levels in assigned macular and peripheral cone and rod precursors at each age.

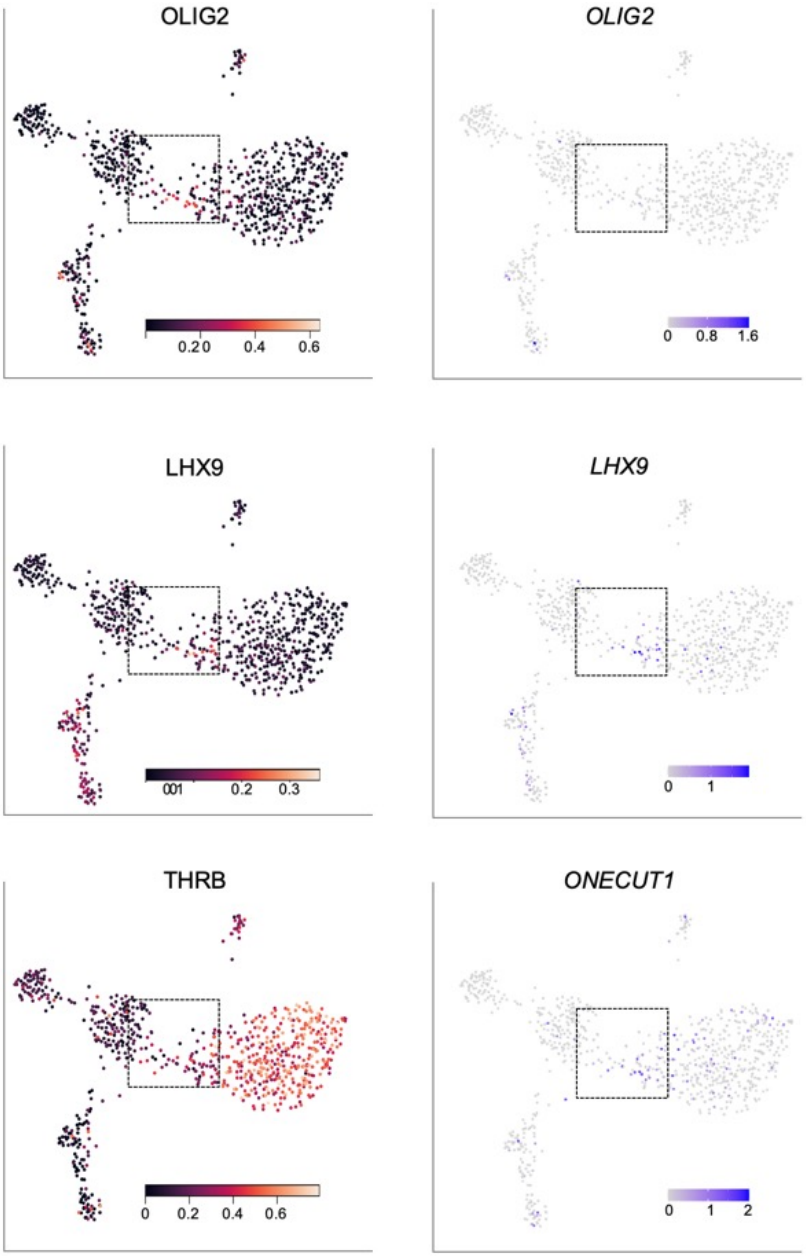


Figure S14: **iCP-enriched regulon activities.** UMAP plots of OLIG2, LHX9, and THRB regulon activities (*left*) and *OLIG2*, *LHX9*, and *ONECUT1* gene expression (*right*), illustrating high OLIG2 regulon activity with minimal *OLIG2* gene expression, concurrent LHX9 regulon activity and *LHX9* gene expression, and *ONECUT1* gene expression in iCP cells preceding increased THRB regulon activity in early LM cones. Boxes indicate bridge region.

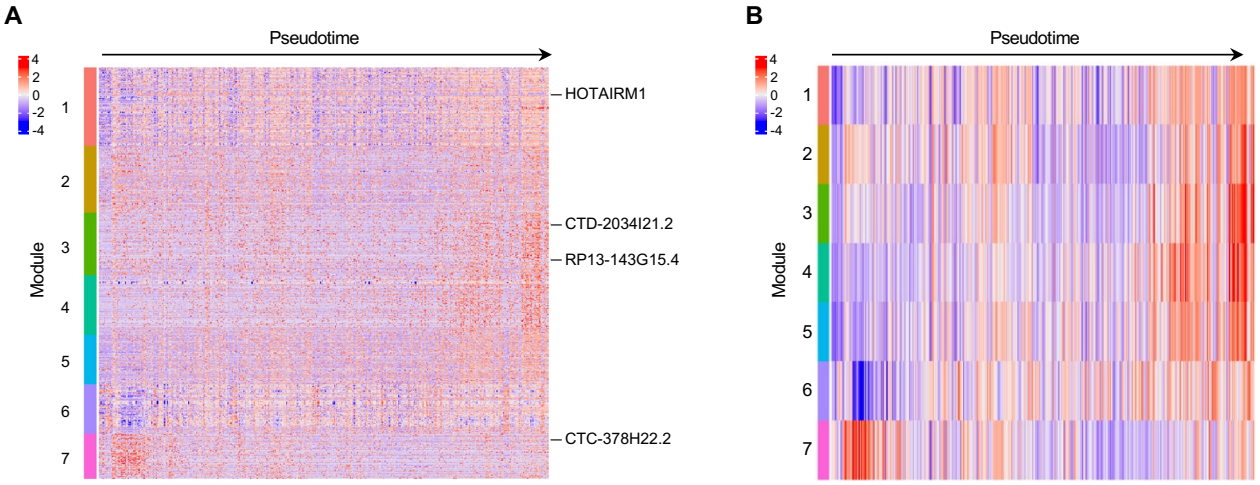


Figure S15: Heatmaps of coregulated gene modules across cone precursor pseudotime. A. Heatmaps of individual genes in each module across the pseudotime trajectory shown in Figure 6C. lncRNA genes of interest are labeled. **B.** Heatmap of averaged gene module expression across pseudotime.

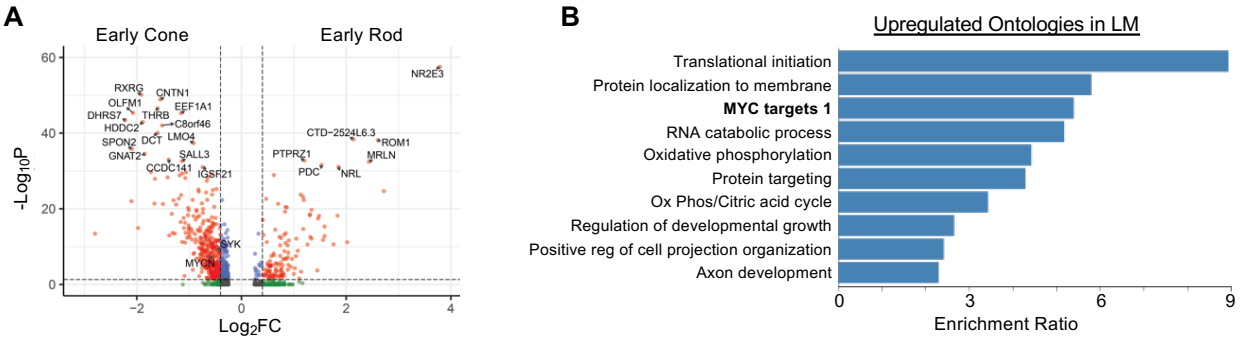


Figure S16: Differential gene expression in early cone and rod precursors. A. Volcano plot of differential expression between early rod cluster ER and cone LM, excluding the 5-cell late-maturing population. pAdj cutoff = 0.05, $\log_2$FC cutoff = |0.4|. Labeled genes: pAdj <10e-32, except for *SYK* and *MYCN.* **B.** Overrepresentation analysis of cone-enriched genes pAdj <0.05, $\log_2$FC ≥ |0.4|.

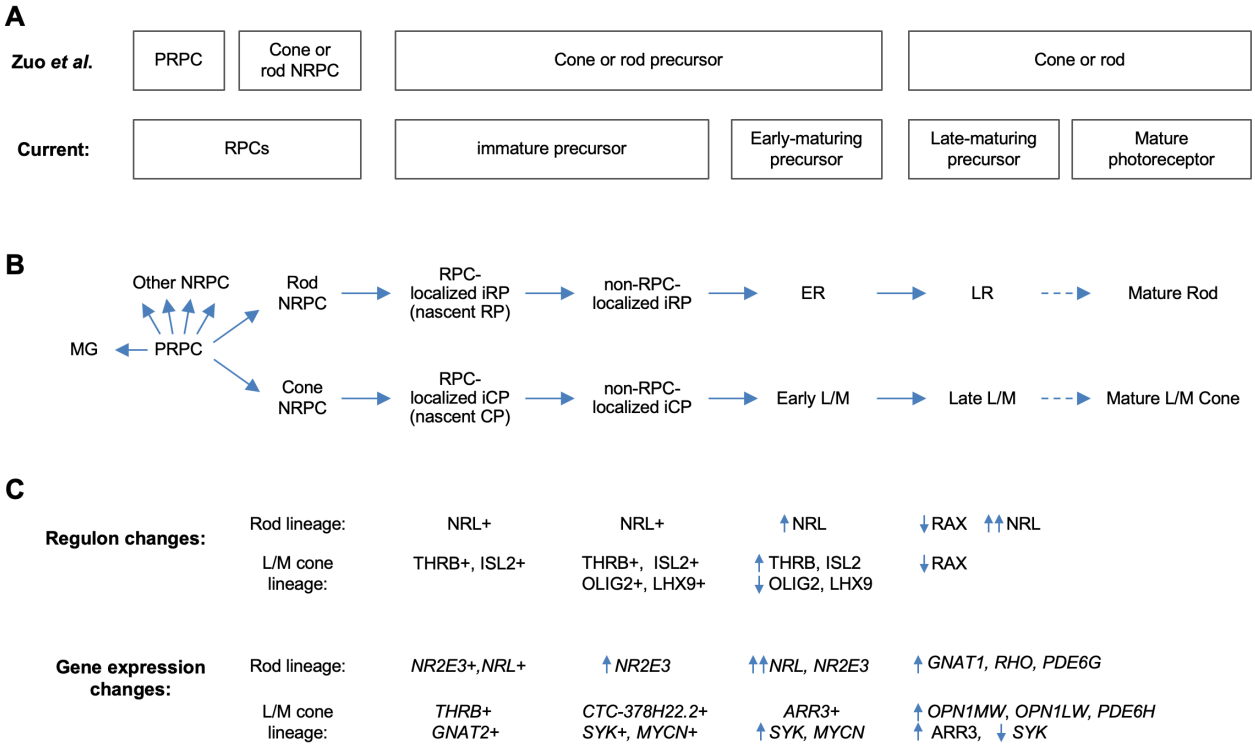


Figure S17: Proposed cell state trajectories in human L/M cone and rod development. A. L/M cone and rod developmental stages discerned in the 3' snRNA-seq analysis of Zuo *et al.*[20] (*top*) and in the current deep, full-length scRNA-seq study (*bottom*). Note discrimination of PRPC and NRPC populations in Zuo *et al.* and discrimination of immature and early-maturing precursors in the current work. Late maturing precursors in fetal retina are hypothesized to be distinguishable from fully mature photoreceptors in postnatal retina. **B**. RPC and L/M cone and rod developmental states referred to in the current study. **C.** Selected cell state features identified in the current work.

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
strain, strain background (*Escherichia coli*)	NEB 10-beta	New England Biolabs	C3019H	Competent cells
cell line (*Homo-sapiens*)	CHLA-VC-RB31	Stachelek *et al*., 2023 DOI: 10.1002/gcc.23120		
cell line (*Mus musculus*)	NIH-3T3	ATCC	CRL-1658.2	.
cell line (*Homo-sapiens*)	HEK-293T	ATCC	CRL-3216	
biological sample (*Human*)	Fetal eyes	Family Planning Associates, Los Angeles, CA		Freshly isolated from fetus
biological sample (*Human*)	Fetal eyes	Advanced Bioscience Resources, Alameda, CA		Freshly isolated from fetus
antibody	anti-NRL (Goat polyclonal)	R&D Systems	CAT# AF2945, RRID: AB_2155098	WB: 1:2000-4000)

antibody	anti-NRL (Mouse monoclonal)	Santa Cruz Biotechnology	CAT# SC-374277, RRID: AB_10991100	WB:1:250, IF:1:50
antibody	anti-RXRγ (Mouse monoclonal)	Santa Cruz Biotechnology	CAT# SC-514134, RRID: AB_2737293	IF: 1:200
antibody	anti-SYK (Mouse monoclonal)	Santa Cruz Biotechnology	CAT# SC1240, RRID: AB_628308	IF:(1:200)
antibody	anti-ARR3 (Rabbit polyclonal)	Li et al., 2003 DOI: 10.1167/iovs.02-0434 Zhang et al., 2001 DOI: 10.1007/978-1-4615-1355-1_33	LUMI-F - hCAR	IF: (1:5000) Cheryl Craft
antibody	anti-RXRγ (Rabbit polyclonal)	Santa Cruz Biotechnology	CAT# SC-555, RRID: AB_2269865	IF: (1:800)
antibody	anti-KI67 (Mouse monoclonal)	BD Bioscience	CAT# 550609, RRID: AB_393778	IF: (1:200)
antibody	anti-GFP and YFP (Goat polyclonal)	Abcam	CAT# ab6673, RRID: AB_305643	IF: (1:500)

antibody	anti-CD133-PE (Mouse monoclonal)	Miltenyi Biotec	CAT# 130-113-108, RRID: AB_2725937	FACS: (1:50)
antibody	Mouse monoclonal anti-CD44-FITC (1:50)	BD Biosciences	CAT# 555478, RRID: AB_395870	FACS: (1:50)
antibody	Mouse monoclonal anti-PNR/NR2E3 (1:50)	R&D Systems	CAT# PP-H7223-00, RRID: AB_2155481	IF: (1:50)
antibody	Mouse monoclonal anti-CD49b-FITC (1:10)	BD Biosciences	CAT# 555498, RRID: AB_395888	FACS: (1:10)
antibody	anti-goat IgG Alexa Fluor 488 (1:300)	Jackson Laboratories	CAT# 705-545-147, RRID: AB_2336933	IF: (1:300)
antibody	anti-mouse IgG Alexa Fluor 680 (Donkey polyclonal)	Life Technologies	CAT# A10038, RRID: AB_2534014	IF: (1:500)
antibody	Donkey polyclonal anti-mouse IgG Alexa Fluor 680 (Donkey polyclonal)	Invitrogen	CAT# A11057, RRID: AB_2534104	IF: (1:500)
recombinant DNA reagent	pLKO.1C-YFP	Lee and Cobrinik, 2020 DOI: 10.2144/btn-2019-0155		GFP version of pLKO.1-Puro

recombinant DNA reagent	pLKO.1C-YFP-shRB1-733	Lee and Cobrinik, 2020 DOI: 10.2144/btn-2019-0155	RRID :Addgene_10878	Pol III based shRNA backbone
recombinant DNA reagent	pcDNA4-His-Max-C-Nrl	Cheng et al., 2004 DOI: 10.1093/hmg/ddh173		Gift from A. Swaroop.
recombinant DNA reagent	pcDNA4-His-Max-C-EF1α-FL-NRL	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses
recombinant DNA reagent	pcDNA4-C-EF1α-FL-NRL	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses
recombinant DNA reagent	pcDNA4-C-EF1α-Tr-NRL	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses
recombinant DNA reagent	pcDNA4-C-EF1α	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses
recombinant DNA reagent	pGL3-SV40	Promega	AT# E1761, RRID:Addgene_173953	
recombinant DNA reagent	pGL3-PDE6B-146	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses
recombinant DNA reagent	pGL3-empty	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses

recombinant DNA reagent	pUltra-EGFP-P2A-Tr-NRL	This paper	Submitted to Addgene, RRID pending	See Materials and Methods NRL isoform analyses
recombinant DNA reagent	pUltra-EGFP	Lou et al., 2012 DOI: 10.1371/journal.pone.0033093	RRID: Addgene_24129	
sequence-based reagent	Gipc1_F	This paper	PCR primers	GGGAAAGGACAAAAGGAACCC
sequenced-based reagent	Gipc1_R	This paper	PCR primers	CAGGGCATTTGCACCCCATGCC
sequenced-based reagent	siRNA: nontargeting control	Thermo Fisher	4390843	Silencer Select
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	1) Del-His F	5'-CCGAAACCATGGCCCTGCCCCCCAGC
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	2) Del-His R	5'-GGGCCATGGTTTCGGAGGCCGTCCG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	3) NRL-no-His F	5'-CCGAAACCATGTCTGTGCGGGAGCTAAACC
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	4) NRL-no-His R	5'-CAGACATGGTTTCGGAGGCCGTCCG

sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	5) pcDNA del NRL F	5'-CCGAAACCGCCGTTCAGAGCACCTTGTGG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	6) pcDNA del NRL R	5'-GAACGGCGGTTTCGGAGGCCGTCCG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	7) PDE -93 F IF	5'-TCTTACGCGTGCTAGAGCGCAGGCCCCCATTTG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	8) PDE +53 R IF	5'-CTTAGATCGCAGATCGGTGGCTGCCTGTCCCTG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	9) pGL3-sv40 SDM F	5'-CTGCGATCAAGCTTGGCATTCCGGTACTG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	10) pGL3-sv40 SDM R	5'-CAAGCTTGATCGCAGATCGGTGGCTG
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	11) pGL3 del SV40 R	5'-CAAGCTTGATCGCAGATCTCGAGCCC
sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	12) Tr-NRL IF pU-G F	5'-GCCTTCTAGAGGATCCATGTCTGTGCGGGAGCTAAACC

sequenced-based reagent	Subcloning PCR primer	Integrated DNA Technologies	13) Tr-NRL IF pU-G R	5'-CGCCGGAGCCGGATCCTCAGAGGAAGAGGTGGGAGGG
sequenced-based reagent	SmartSeq library amplification primer	Integrated DNA Technologies	5' PCR Primer II A	5'-AAGCAGTGGTATCAACGCAGAGT
sequenced-based reagent	RACE primer	Integrated DNA Technologies	5' RACE primer for SmartSeq cDNA	5'-AAGCAGTGGTATCAACGCAGAGTACGGG
sequenced-based reagent	NRL reverse primer for 5' RACE	Integrated DNA Technologies	Rev NRL-Full ex3	5'-GGTTTAGCTCCCGCACAGACATCGAGAC
sequenced-based reagent	RXRG reverse primer for 5' RACE	Integrated DNA Technologies	Rev RXR ex5	5'-GAAGAACCCTTTGCAGCCTTCACAACTG
sequenced-based reagent	RACE primer	Integrated DNA Technologies	3' RACE primer for SmartSeq cDNA	5'-AAGCAGTGGTATCAACGCAGAGTACTTTT
sequenced-based reagent	THRB1 forward primer for 3' RACE	Integrated DNA Technologies	Forw TRb1 ex4	5'-GCCTTACAGCCTGGGACAAACC
sequenced-based reagent	THRB2 forward primer for 3' RACE	Integrated DNA Technologies	Forw TRb2 ex1	5'-CCCTGGAAACATGTTTAAAGCAAGGACT

sequenced-based reagent	*In situ* hybridization chain reaction probes	Molecular Instruments Inc.	FL-NRL exon 1, 2	Target sequence Ensembl ID: ENST00000 619224.1
sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*CHRNA1*	Target sequence NCBI accession number: NM_000079.4
sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*RP13-143G15.4*	
sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*CTC-378H22.2*	Target sequence Ensembl ID: ENST00000 559786.1
sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*HOTAIRM1*	Target sequence NCBI accession number: NR_038366.1
sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*CTD-2034I21.2*	Target sequence NCBI accession number: XR_001752169.1
sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*GNAT2*	Target sequence NCBI accession number: NM_005272.5

sequenced-based reagent	hybridization chain reaction probes	Molecular Instruments Inc.	*NR2E3*	Target sequence NCBI accession number: NM_01249.4
commercial assay or kit	In-Fusion HD Cloning	Clontech	Clontech:639647	
commercial assay or kit	SMARTer Ultra Low RNA Kit for the Fluidigm C1 System	Clontech (Takara Bio)	CAT# 634835/634935	
commercial assay or kit	SMART-Seq V4 Ultra Low Input RNA Kit	Takara Bio	CAT# 63891	
commercial assay or kit	Nextera XT DNA Library Preparation Kit	Illumina	CAT# FC-131-1096	
commercial assay or kit	CloneAmp HiFi PCR Premix	Clontech (Takara Bio)	Takara Bio #639298	
commercial assay or kit	Quant-iT PicoGreen dsDNA Assay Kit	Life Technologies	CAT# P11496	
commercial assay or kit	Nano-Glo Dual-Luciferase Reporter Assay System	Promega, Inc	CAT# N1610	
chemical compound, drug	GS-9876 (SYK inhibitor)	MedChemExpress	CAT# HY-109091	
software, algorithm	TrimGalore	Felix Krueger GitHub		https://github.com/FelixKr

				ueger/TrimGalore
software, algorithm	Cutadapt	Martin, 2011 DOI: 10.14806/ej.17.1.200		https://cutadapt.readthedocs.io/en/stable/
software, algorithm	HISAT2	Kim et al., 2019 DOI: 10.1038/s41587-019-0201-4		https://github.com/DaehwanKimLab/hisat2
software, algorithm	StringTie	Pertea et al., 2015 DOI: 10.1038/nbt.3122		minmmhttp://ccb.jhu.edu/software/stringtie/
software, algorithm	Snakemake/ARMOR	Orjuela et al., 2019 DOI: 10.1038/nbt.3122		https://github.com/whtns/ARMOR
software, algorithm	Minimap2	Li, 2018 DOI: 10.1093/bioinformatics/bty191		https://lh3.github.io/minimap2/
software, algorithm	WebGestalt	Liao et al., 2019 DOI: 10.1093/nar/gkz401		http://www.webgestalt.org/
software, algorithm	pySCENIC	Van de Sande et al., 2020 DOI: 10.1038/s41596-020-0336-2		https://github.com/aertslab/pySCENIC

software, algorithm	Tximport	Soneson et al., 2015 DOI: 10.12688/f1000research.7563.1		https://bioconductor.org/packages/release/bioc/html/tximport.html
software, algorithm	Seurat v3 (full-length scRNA-seq), Seurat v5 (3' snRNA-seq)	Stuart et al., 2019 DOI: 10.1016/j.cell.2019.05.031		https://satijalab.org/seurat/index.html
software, algorithm	EnhancedVolcano	Blighe et al., 2018 Github		https://github.com/kevinblighe/EnhancedVolcano
software, algorithm	genesorteR (v0.4.3)	Ibrahim and Kramann, 2019 DOI: 10.1101/676379		http://github.com/mahmoudibrahim/genesorteR
software, algorithm	Wiggleplotr (v1.13.1)	Alasoo, 2019 Bioconductor package		https://github.com/kaurala soo/wiggleplotr
software, algorithm	DEXSeq	Anders et al., 2012 DOI: 10.1101/gr.133744.111		https://bioconductor.org/packages/release/bioc/html/DEXSeq.html
software, algorithm	velocytoR	La Manno et al., 2018 DOI: 10.1038/s41586-018-0414-6		https://velocyto.org/
software, algorithm	Monocle3	Cao et al., 2019 DOI: 10.1038/s415		https://cole-trapnell-lab.github.io/monocle3/

		86-019-0969-x		
software, algorithm	QuPath	Bankhead et al., 2017 DOI: 10.1038/s41586-019-0969-x		https://qupath.github.io/
software, algorithm	StarDist (QuPath implementation)	Schmidt et al., 2018 DOI: 10.1038/s41586-019-0969-x		https://qupath.readthedocs.io/en/stable/docs/advanced/stardist.html