Single cell transcriptomics reveals early photoreceptor states, cell-specific transcript isoforms, and cancer-predisposing features

Short: Early human photoreceptor precursor cell states

Dominic W.H. Shayler,[1,2] Kevin Stachelek,[1,3] Linda Cambier,[1] Sunhye Lee,[1] Jinlun Bai, [1,2] Mark W. Reid, [1] Daniel J. Weisenberger,[4,5] Bhavana Bhat,[1] 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,[4,9] Cheryl Mae Craft,[10,11] Hardeep P. Singh,[1,9] David Cobrinik[1,4,5,9]*

[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 Biochemistry & Molecular Medicine, 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. Here, we used deep full-length single-cell RNA-sequencing to distinguish post-mitotic cone and rod developmental states and cone-specific features that contribute to retinoblastomagenesis. The analyses revealed early post-mitotic cone- and rod-directed populations characterized by higher THRB or NRL regulon activities, an immature photoreceptor precursor population with concurrent cone and rod gene and regulon expression, and distinct early and late cone and rod maturation states distinguished by maturation-associated declines in RAX regulon activity. Unexpectedly, both L/M cone and rod precursors co-expressed *NRL* and *THRB* RNAs, yet they differentially expressed functionally antagonistic *NRL* isoforms and prematurely terminated *THRB* transcripts. Early L/M cone precursors exhibited successive expression of lncRNAs along with *MYCN*, which composed the seventh most L/M-cone-specific regulon, and *SYK*, which contributed to the early cone precursors' proliferative response to *RB1* loss. These findings reveal previously unrecognized photoreceptor precursor states and a role for early cone-precursor-intrinsic *SYK* expression in retinoblastoma initiation.

Key words: cone photoreceptor, rod photoreceptor, photoreceptor development, retinoblastoma, scRNA-seq, regulon, lncRNA, trajectory

Introduction

The vertebrate retina forms from optic vesicle retinal progenitor cells (RPCs), which give rise to seven major post-mitotic cell types including early born retinal ganglion cells, cone photoreceptors, horizontal cells, and amacrine cells and later born rod photoreceptors, bipolar cells, and Müller glia. The orderly generation of different retinal cell types is governed by the combined actions of a series of transcription factors that enable progressive RPC lineage restriction, fate determination, and post-mitotic maturation (*1*, *2*). However, despite recent advances (*1*), major events underlying retinal cell fate determination and maturation remain unclear. For example, during photoreceptor development, it is unclear if fate is fully determined in RPCs, where OTX2 and ONECUT1 are thought to control the post-mitotic expression of long- or medium-wavelength (L/M) cone fate determinant TRβ2 and rod determinant NRL (*3*), or is fate determined post-mitotically in cells with TRβ2 and NRL co-expression (*4*). Following photoreceptor fate commitment, post-mitotic developmental stages have been defined based on morphologic features and phototransduction-related protein or gene 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. Resolving these issues may provide insight into photoreceptor developmental abnormalities underlying retinal dystrophies, retinal degeneration, or the retinal cone precursor-derived cancer, retinoblastoma (*7*, *8*).

While much understanding of retinal development derives from studies of model organisms (*2*, *9*), most models do not recapitulate human retinal features such as a rod-free fovea or tri-color vision. Another human-specific feature is the maturing cone precursors' proliferative response to loss of the retinoblastoma protein (pRB), representing the first step in retinoblastoma tumorigenesis (*10*, *11*). While the proliferative response depends on the human cone precursors' high MDM2 and MYCN expression (*10*), the developmental basis and critical features of this state have not been defined. Due to the potential importance of species-specific

traits, analysis of human tissue is required to fully define disease-relevant photoreceptor features.

Transcriptome profiling has provided numerous insights into retinal gene expression and cell state changes. Bulk profiling has revealed overall temporal patterns of gene expression in mouse (*12*) as well as human retinae (*6*, *13*). Single-cell RNA-sequencing (scRNA-seq) has further revealed cell type-specific features of photoreceptor fate specification and maturation. scRNA-seq studies employing 3' end-counting have defined age-related post-mitotic transition populations, fate-determining features of post-mitotic photoreceptor precursors (*14–16*), and gene expression changes associated with the cone fate decision and early development (*14–19*). 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 cell states in individual cells. Moreover, while several photoreceptor related transcription factors have been identified, the factors that govern human photoreceptor developmental states have not been examined in an unbiased manner.

In this study, we sought to further define the transcriptomic underpinnings of human photoreceptor developmental trajectories. We generated deep, full-length single-cell transcriptomes of human retinal progenitor cells (RPCs) and developing photoreceptors from fetal week 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 unrecognized photoreceptor developmental states, identified cell type-specific expression of photoreceptor fate-determining transcript isoforms, elucidated post-mitotic developmental trajectories, and defined retinoblastoma cell-of-origin features that 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 fetal week 13–19 retinae 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. Exclusion of all cells with <100,000 reads and 18 cells expressing markers of retinal ganglion, amacrine, and/or horizontal cells (*POU4F1*, *POU4F2*, *POU4F3*, *TFAP2A*, *TFAP2B*, *ISL1*) and lacking photoreceptor lineage marker *OTX2* 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, isolation methods, and sequencing runs (Figure 1B, S1D-F).

Low resolution Louvain clustering 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 (Figure 1C, S1G-J). In UMAP space, the cluster designated immature photoreceptor precursors (iPRPs) intermixed with the RPC/MG population, extended towards early cone precursors, and was predominantly comprised of cells expressing the L/M cone determinant *THRB* (Figures 1B,C).

Two clusters 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* (Figure 1C, S1I).

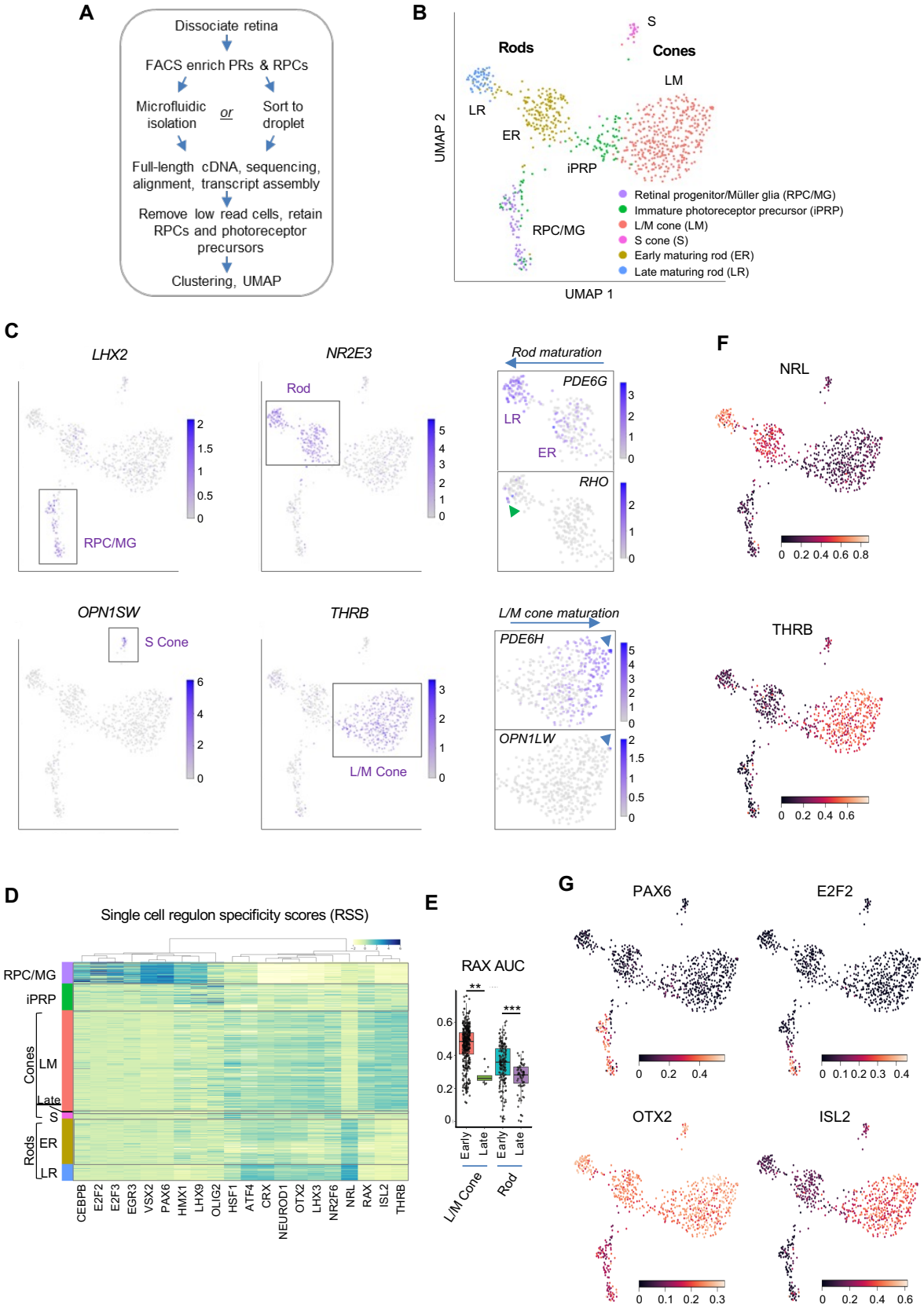

A
Dissociate retina
FACS enrich PRs & RPCs
Microfluidic isolation
or
Sort to droplet
Full-length cDNA, sequencing, alignment, transcript assembly
Remove low read cells, retain RPCs and photoreceptor precursors
Clustering, UMAP
B
Rods
Cones
S
LM
LR
ER
iPRP
RPC/MG
UMAP 2
UMAP 1
Retinal progenitor/Müller glia (RPC/MG)
Immature photoreceptor precursor (iPRP)
L/M cone (LM)
S cone (S)
Early maturing rod (ER)
Late maturing rod (LR)
C
LHX2
RPC/MG
NR2E3
Rod
Rod maturation
PDE6G
LR
ER
RHO
OPN1SW
S Cone
THRB
L/M Cone
L/M cone maturation
PDE6H
OPN1LW
F
NRL
THRB
D
Single cell regulon specificity scores (RSS)
RPC/MG
iPRP
Cones
LM
Late
S
Rods
ER
LR
CEBPB
E2F2
E2F3
EGR3
VSX2
PAX6
HMX1
LHX9
OLIG2
HSF1
ATF4
CRX
NEUROD1
OTX2
LHX3
NR2F6
NRL
RAX
ISL2
THRB
E
RAX AUC
Early
Late
L/M Cone
Rod
G
PAX6
E2F2
OTX2
ISL2

Figure 1: Photoreceptor-enriched full-length scRNA-seq identifies regulon-defined RPC and photoreceptor precursor states. A. Overview of sample collection and sequencing. RPCs and developing photoreceptors were FACS-enriched and isolated directly or using C1 microfluidics. **B.** UMAP plot of final dataset, low resolution clusters labeled by cell type. **C.** Expression of marker genes for RPC/MGs (*LHX2)*, rods (*NR2E3)*, S cones (*OPN1SW)*, L/M cones (*THRB*), rod maturation (*PDE6G*, *RHO*) and cone maturation (*PDE6H*, *OPN1LW*). Arrowheads: Late maturing *RHO*+ rods (*top*), late maturing *OPN1LW*+ cones (*bottom*). **D.** Ward-clustered heatmap of the highest scoring SCENIC regulons in each cluster, displaying Z-score normalized regulon activities. Late = late maturing L/M cones. **E.** 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. **F,G.** UMAP plots of regulon AUC values for (F) NRL and THRB (rod and L/M cone) and (G) PAX6, E2F2, OTX2, ISL2 (MG/RPC, RPC, PR, L/M cone). See also Figure S1.

Differential expression analysis revealed other genes upregulated in late rods (*GNB1, SAMD7*, *NT5E*) (*20*, *21*) as well as the downregulated *CRABP2*, *DCT*, and *FABP7* (Figure S1K, Table S1). Genes upregulated in the LR cluster were enriched for photoreceptor and light sensing gene ontologies (Figure S1L) including spectrin binding, likely relating to proteins that control photoreceptor polarity and synapse formation (*22*, *23*), and purine biosynthesis and ribonucleotide metabolic processes, potentially related to the developing photoreceptors' high NAD+ requirements (*24*).

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*) (*25–27*)**,** novel S cone enriched genes (*MEGF10, NRXN3*, *ACKR3*), and the L/M cone transcription factor *ISL2* (*15*), among others (Figure 1B,C, S1M, 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 S1N) 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 363-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 1C, S1O,P), analogous to *RHO*, *PDE6G*, and *GNGT1* upregulation in late rods. Compared to the early L/M population, late L/M cones had upregulation of the three M-opsin genes (*OPN1MW*, *OPN1MW2*, *OPN1MW3*), *TTR*, encoding the retinol-binding protein transthyretin, *PCP4*, a small protein that binds calmodulin previously noted in foveal cones (*28*), and *MYL4*, a myosin light chain gene upregulated in retinal organoid L/M cones (*27*), among others (Figure S1O, 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*).

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 (*29*). 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 1D-G, Table S4). SCENIC also distinguished ER and LR states via the latter's increased NRL, CRX, ATF4, and LHX3 and decreased HMX1 and RAX regulon activities ($p < 0.0005$ for each, Dunn test) (Figure 1D). RAX activity also decreased in the 5-cell late maturing L/M cone group (Figure 1E), 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 1D, Table S4). 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 (*3*). Thus, deep, full-length scRNA-seq enabled identification of regulons underlying RPC and developing photoreceptor states at the single cell level.

Novel *NRL* and *THRB* isoforms in rod and cone populations

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 populations, with mean *NRL* expression only 3.8-fold higher in the ER vs LM cluster, mean *THRB* expression only 2.5-fold higher in LM vs ER, and mean *RXRG* expression 4.7-fold higher in LM vs ER (Figure S1I, J). Cone *NRL* expression was

unexpected given NRL's role in rod differentiation (*27*) and the rod-specific NRL regulon activity (Figure 1D,F). Similarly, rod *RXRG* and *THRB* expression were unexpected given their roles in cone gene expression and differentiation (*30*, *31*) and the L/M cone-specific THRB regulon activity (Figure 1D,F). Accordingly, we used our full-length scRNA-seq data to determine if cell type-specific regulation of transcript isoforms underlies the far higher NRL protein in rods and TRβ2 and RXRγ proteins in cones by interrogating inferred Ensembl isoforms, exon coverage, and targeted long read cDNA sequencing.

For *NRL*, the most highly assigned transcript isoform (*ENST00000397002*) encoded the canonical full-length NRL (FL-NRL) (RefSeq NP_001341697.1), while the second-most assigned isoform (*ENST00000560550*) encoded a previously uncharacterized transcript that used an alternative "P2" promoter and first exon, here termed exon 1T (Figure 2A-C). The novel transcript was predicted to encode an N-terminally truncated NRL (Tr-NRL) retaining the leucine zipper DNA binding domain but lacking the minimal transactivation domain (*32*) (Figure 2C). While both isoforms were inferred to be more highly expressed in the ER vs LM cluster, the average FL-NRL (*ENST00000397002*) : Tr-NRL (*ENST00000560550*) ratio was 2.8:1 in early rods and 0.72:1 in L/M cones (Figure 2A,B). Consistent with the assigned isoform ratios, mean Tr-NRL exon 1T coverage was higher in S and L/M cones while FL-NRL exon 1 coverage was higher in rods (Figure 2C, 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 2D). 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 (Figure 2E,F).

While the Tr-NRL isoform was not 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 2C), was previously identified in adult retina (*33*). Concordantly, we detected reads spanning the unique DD10 splice junction, yet at lower levels than the unique

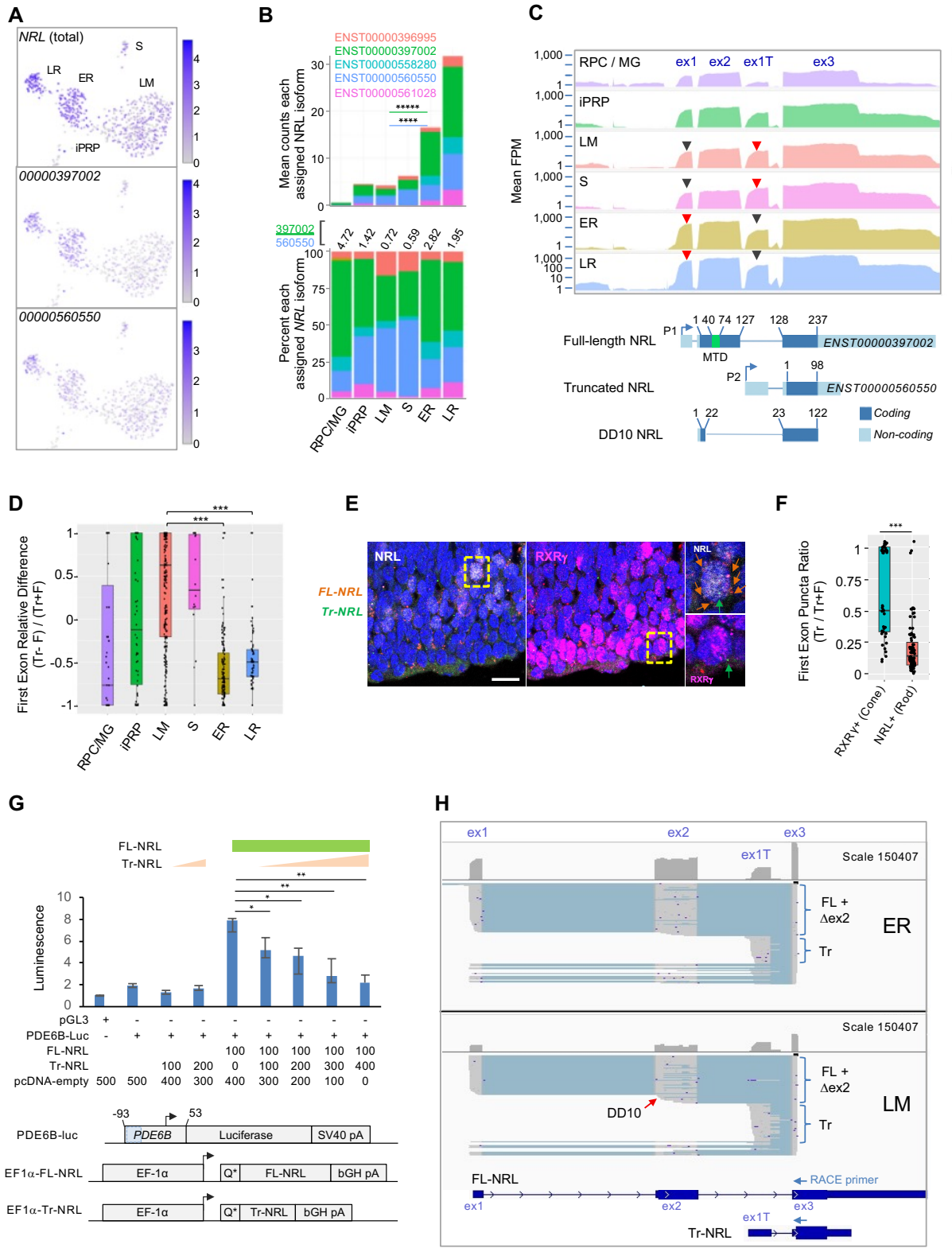

A
NRL (total)
S
LR
ER
LM
iPRP
00000397002
00000560550
B
ENST00000396995
ENST00000397002
ENST00000558280
ENST00000560550
ENST00000561028
Mean counts each assigned NRL isoform
397002
560550
4.72
1.42
0.72
0.59
2.82
1.95
Percent each assigned NRL isoform
RPC/MG
iPRP
LM
S
ER
LR
C
RPC / MG
ex1
ex2
ex1T
ex3
iPRP
LM
S
ER
LR
Mean FPM
Full-length NRL
P1
MTD
ENST00000397002
Truncated NRL
P2
ENST00000560550
DD10 NRL
Coding
Non-coding
D
First Exon Relative Difference (Tr- F) / (Tr+F)
RPC/MG
iPRP
LM
S
ER
LR
E
NRL
RXRγ
FL-NRL
Tr-NRL
F
First Exon Puncta Ratio (Tr / Tr+F)
RXRγ+ (Cone)
NRL+ (Rod)
G
FL-NRL
Tr-NRL
Luminescence
pGL3
PDE6B-Luc
FL-NRL
Tr-NRL
pcDNA-empty
PDE6B-luc
PDE6B
Luciferase
SV40 pA
EF1α-FL-NRL
EF-1α
Q*
FL-NRL
bGH pA
EF1α-Tr-NRL
Tr-NRL
H
ex1
ex2
ex3
ex1T
Scale 150407
FL + Δex2
Tr
ER
LM
DD10
FL-NRL
RACE primer
Tr-NRL

Figure 2: Differential expression of *NRL* isoforms in rod and cone populations. A. Expression of *NRL* gene and the most highly assigned Ensembl isoforms *ENST00000397002*, encoding full-length NRL (FL-NRL) and *ENST00000560550*, encoding truncated NRL (Tr-NRL). **B.** Mean *NRL* isoform assignments for clusters defined in Figure 1B, presented as normalized transcript counts (*top*) and percentage of total counts (*bottom*). Significance lines for LM vs. ER fold change, colored by isoform. ****, p <0.0002; *****, <0.000001 (bootstrapped Welch's t-test). **C.** *Top*: Mean read counts across Ensembl annotated *NRL* exons for each cluster. *Bottom*: Transcript structures for *ENST00000397002, ENST00000560550,* and DD10, 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.** Combined NRL and RXRγ immunohistologic staining and RNA FISH with probes specific to truncated Tr-NRL exon 1T (green puncta) and to FL-NRL exons 1 and 2 (orange puncta) in FW16 retina. Boxed regions enlarged at right show a RXR^{lo}, NRL^{+} rod with one Tr-NRL and six FL-NRL puncta (*top*) and an $RXR\gamma^{hi}$, NRL^{-} cone with one Tr-NRL and no FL-NRL puncta (*bottom*), as indicated with same-color arrows. Scale bar: 10 μm. **F.** Ratio of fluorescent puncta observed in experiment depicted in (E) for NRL+ or $RXR\gamma^{hi}$ cells where Tr puncta >0. ***; p <0.0005 (Welch's t-test). **G.** *Top*: PDE6B-luciferase reporter activity in NIH3T3 cells transfected with indicated amounts (ng) of pcDNA4C-EF1α and derived FL-NRL and Tr-NRL. *Bottom:* Expression constructs. Error bars = standard deviation of triplicate measurements. *, p <0.05; **, <0.005 (Student's t-test). Data representative of two experiments in NIH3T3 and one in 293T. **H.** 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 transcript structures (*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.

Tr-NRL junction (5,942 versus 57,048). As both Tr-NRL and DD10 lack the NRL transactivation domain and as DD10 interferes with FL-NRL transactivation (*34*), we examined if Tr-NRL similarly opposes FL-NRL transcriptional activity. We found that in luciferase assays, Tr-NRL suppressed FL-NRL activation of a *PDE6B* promoter (Figure 2G), demonstrating that the cone-enriched Tr-NRL antagonized FL-NRL-mediated transactivation.

To further define *NRL* isoforms, we performed NRL 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 2H, Figure S2A). 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 2H, Figure S2A). 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 far 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 in early rods and disrupted FL-NRL and Tr-NRL isoforms in LM 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 (*35*), which expressed FL-NRL- and Tr-NRL transcripts in a cone-like 0.57:1 ratio, with an antibody that recognizes both FL- and Tr-NRL proteins (Figure S2B,C). As in other retinoblastoma cells (*36*), FL-NRL increased in response to retinoic acid and proteasome inhibition, whereas endogenous Tr-NRL was not detected (Figure S2D). 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 S2E-H). These findings suggest that Tr-NRL is poorly translated or unstable in most cone precursors.

For *THRB*, the most highly assigned transcript isoforms included the canonical L/M cone-specific TRβ2 (*ENST00000280696)* and TRβ1 (*ENST00000396671*) (Figure 3A-C). The different *THRB* exons were similarly represented in iPRP and LM clusters, whereas LR cells, with less overall *THRB* expression, preferentially expressed *THRB* exons 1-6 and the *THRB* 3' untranslated region (UTR) (Figure 3B,C). Notably, a higher percentage of reads extended from the *THRB* exon 4 and exon 6 splice donor sequences into the subsequent introns in LR versus LM cells (Figure 3D-E) ($p \leq 0.001$ for both, two-tailed Chi square test), suggesting that premature transcription termination (PTT) in introns 4 and 6 limits *THRB* expression.
To further evaluate *THRB* isoforms, we used *THRB* 3' RACE and long-read sequencing on pooled single cell cDNA libraries from 23 L/M cone cells and five late rod (LR) cells. RACE reactions were performed separately with forward primers complementary to the TRβ1-specific exon 4 and to the TRβ2-specific first exon (Figure 3F). Sequencing of LM and LR RACE products corroborated pronounced PTT in introns 4 and 6, with greater PTT in intron 6 in LR versus LM cells (Figure 3F, Figure S3). However, we did not corroborate the higher PTT in intron 4, 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 solely in TRβ2 transcripts (Figure 3F, S3A). 3' RACE transcripts rarely extended into the 3' UTR in LM or LR cells, suggesting that such sequences are distinct from coding transcripts and do not reflect protein-coding RNA expression. Together, 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 3' UTR expression.

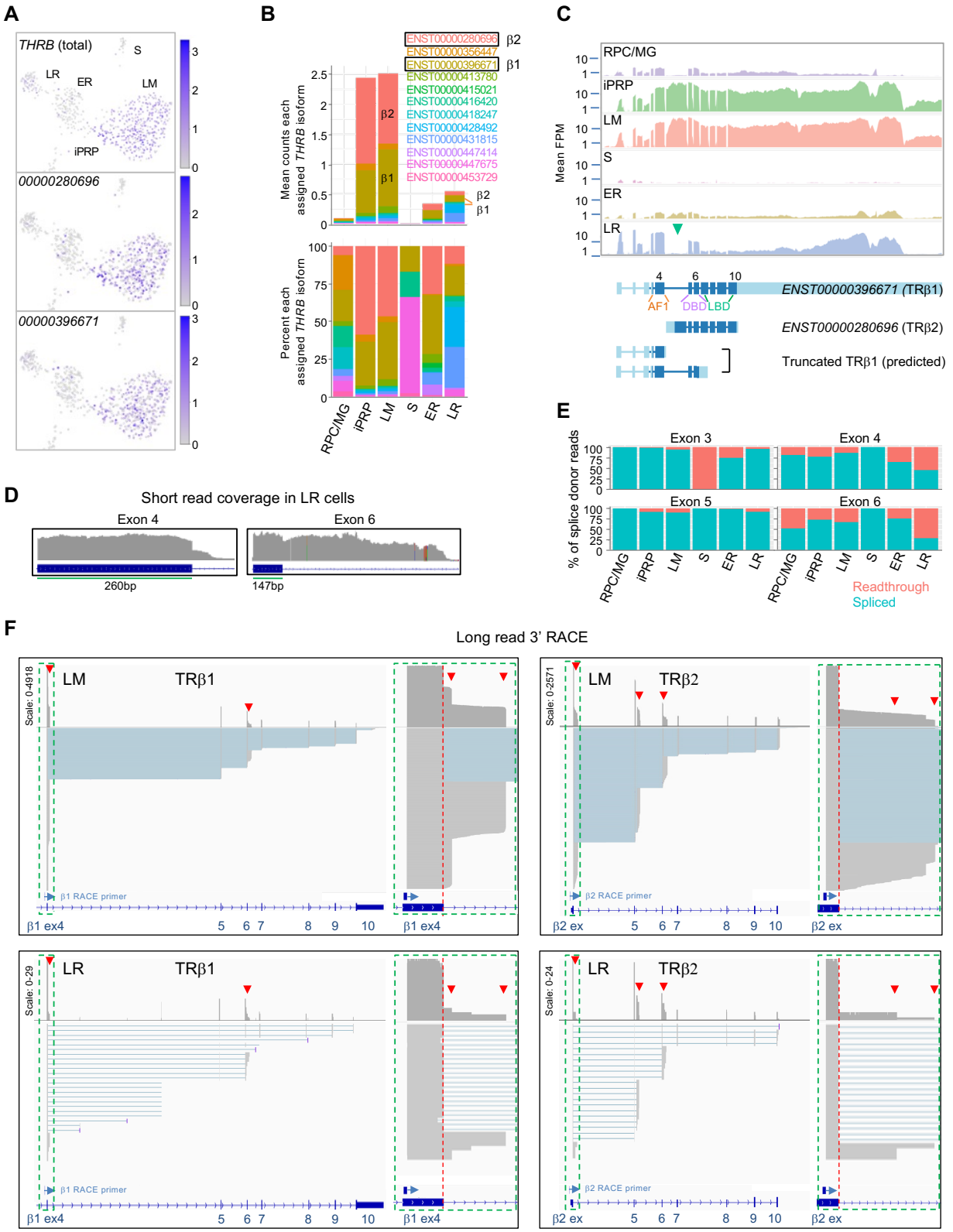

A
THRB (total)
S
LR
ER
LM
iPRP
00000280696
00000396671
B
Mean counts each assigned THRB isoform
Percent each assigned THRB isoform
ENST00000280696 β2
ENST00000356447
ENST00000396671 β1
ENST00000413780
ENST00000415021
ENST00000416420
ENST00000418247
ENST00000428492
ENST00000431815
ENST00000447414
ENST00000447675
ENST00000453729
RPC/MG
iPRP
LM
S
ER
LR
C
Mean FPM
ENST00000396671 (TRβ1)
AF1
DBD
LBD
ENST00000280696 (TRβ2)
Truncated TRβ1 (predicted)
D
Short read coverage in LR cells
Exon 4
Exon 6
260bp
147bp
E
% of splice donor reads
Exon 3
Exon 4
Exon 5
Exon 6
Readthrough
Spliced
F
Long read 3' RACE
LM
TRβ1
TRβ2
LR
β1 RACE primer
β2 RACE primer
β1 ex4
β2 ex

Figure 3: Premature transcription termination of *THRB* isoforms in rod and cone populations. A. Expression of *THRB* and two highly assigned isoforms (*ENST00000280696* and *ENST00000396671*) in LM and LR populations. **B.** *THRB* isoform assignments for low resolution clusters presented as total transcript counts (*top*) and as a percentage of the total counts (*bottom*). **C.** *Top*: Mean read counts across Ensembl annotated *THRB* exons. *Bottom*: Transcript structures for *ENST00000396671* (TRβ1), *ENST00000280696* (TRβ2), and predicted TRβ1 truncations. Green arrowhead: First TRβ2 exon. Exon numbers[84] are indicated above and protein domains (AF1, DNA-binding (DBD), and ligand binding (LBD)) below TRβ1. **D.** Compiled read coverage for LR cells across *THRB* exons 4 and 6 splice donor site. **E.** Percentage of exon 3-6 splice donor region reads that are spliced and the exon splice donor or readthrough to the subsequent intron. **F.** Long-read nanopore sequencing of pooled 3' RACE reactions initiated with *THRB* TRβ1 exon 4 (*left*) or TRβ2 exon 1 (right) primers and performed on cDNA libraries from 21 LM cells (*top*) and 5 LR cells (*bottom*). Each schematic shows total exon coverage (*above*) and individual transcript structures (*below)*, where expressed sequences are grey and introns light blue. Each panel shows the entire TRβ1 or TRβ2 locus with first exons (in green boxes) shown in detail at right. Red arrowheads: intronic premature transcription termination. Ensembl TRβ1 and TRβ2 transcript isoforms and RACE primer positions are shown below.

For *RXRG*, short read sequencing reads were assigned to several isoforms that differed in their 5' promoter position and exon utilization (Figure S4A-C). However, 5' RACE and long-read sequencing did not support differential isoform expression (Figure S4D), and quantitative imaging revealed an average ~3.5-fold higher RXRγ protein in cones compared to rods (Figure S4E). Thus, long-read sequencing clarified that *RXRG* expression is only moderately higher in human cone versus rod precursors without evidence of cell-type specific isoforms.

Two post-mitotic photoreceptor precursor populations

To further define photoreceptor developmental states, we subdivided the initial six cell populations with higher resolution clustering, identified high resolution cluster-associated genes and regulons, and inferred each cell's rate and direction of transcriptomic change using RNA velocity (*37*) (Figure 4A,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 yielded a new cluster, here designated transitional rods (TR), derived from the low resolution iPRP and rod clusters (Figure 4A,C, S5A). Similar clusters were observed at reduced k.param values that defines nearest neighbors.

The distinction between RPCs and MGs 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 4D, S5A,B). 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* (*12*, *38*, *39*) (Figure 4C,D, S5A). 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 S5B,C), consistent with early MGs resembling quiescent RPCs (*40*).

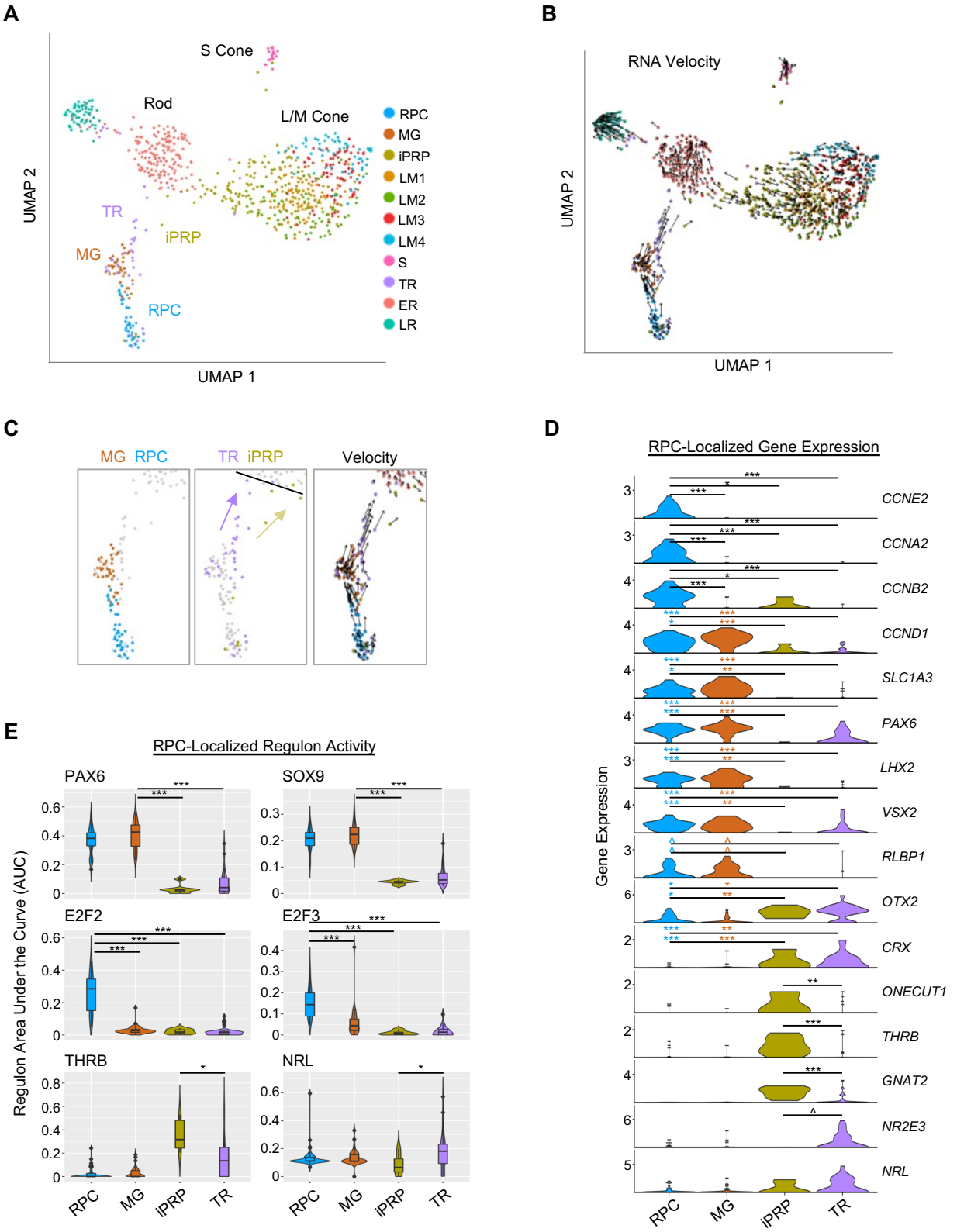

A
S Cone
Rod
L/M Cone
TR
iPRP
MG
RPC
UMAP 2
UMAP 1
RPC
MG
iPRP
LM1
LM2
LM3
LM4
S
TR
ER
LR
B
RNA Velocity
UMAP 2
UMAP 1
C
MG RPC
TR iPRP
Velocity
D
RPC-Localized Gene Expression
Gene Expression
CCNE2
CCNA2
CCNB2
CCND1
SLC1A3
PAX6
LHX2
VSX2
RLBP1
OTX2
CRX
ONECUT1
THRB
GNAT2
NR2E3
NRL
RPC
MG
iPRP
TR
E
RPC-Localized Regulon Activity
Regulon Area Under the Curve (AUC)
PAX6
SOX9
E2F2
E2F3
THRB
NRL
RPC
MG
iPRP
TR

Figure 4: Two post-mitotic photoreceptor precursor populations rapidly expressing rod or cone markers. A. UMAP plot colored by high resolution clusters. **B.** RNA Velocity plots with cell clusters as in A and cell velocity vectors. **C.** Enlarged view highlighting RPC and MG clusters (*left*), RPC-localized iPRP and TR clusters (*middle*), and RNA velocity (*right*). Black line: limit of the RPC-localized region. Arrows depict inferred trajectories. **D.** Violin plots depicting expression of selected genes in RPC, MG, and RPC-localized iPRP and TR 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. ^, $p < 0.1$; *, < 0.05; **, < 0.005; ***, < 0.0005 (post-hoc Dunn test).

High resolution clustering also resolved iPRP and TR cells positioned near the RPCs in UMAP space (Figure 4C). RNA velocity suggested that these 'RPC localized' cell groups flowed towards the larger iPRP population and towards early maturing rods, respectively (Figure 4B). Compared to RPCs and MG, RPC-localized iPRP and TR 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 4D), consistent with their being immediately post-mitotic photoreceptor precursors. The iPRP cells in this region also upregulated the early cell fate determinant *ONECUT1* (*41*, *42*), the L/M cone determinant *THRB*, and the cone differentiation marker *GNAT2,* but lacked the rod determinant *NR2E3*, indicative of a cone-bias (Figure 4D). In contrast, TR cells in the same region trended towards higher expression of the rod-related *NRL* and the rod-specific *NR2E3* (p= 0.19, p=0.054, respectively). Prior droplet scRNA-seq analyses did not resolve early post-mitotic cone versus rod precursors (*15*, *16*), nor could we discriminate early post-mitotic *THRB*+ iPRP and *NR2E3*+ TR populations through a similar analysis of single cell transcriptomes generated by 3' end-counting. These findings suggest that deep sequencing enabled detection of *THRB*, *NR2E3*, and perhaps other low-expressed RNAs that are important to discrimination of early iPRP and TR populations.

SCENIC regulons further clarified the identities of the four RPC-localized clusters (Figure 4E, S5D). The MG cluster was best specified (*i.e.*, had highest regulon specificity scores) by PAX6 and SOX9, both described in RPCs and MGs (*38*, *43*, *44*), whereas RPCs were best specified by E2F2 and E2F3. These RPC and MG regulons were low or absent in the early post-mitotic iPRP and TR cells, suggesting they had undergone an abrupt cell state change. Moreover, immediately post-mitotic iPRPs and TRs that colocalized with RPCs showed differential *THRB* and *NRL* regulon activities. iPRP cells had greater *THRB* regulon signal than TR cells and TR cells had higher NRL regulon signal than iPRPs ($p<0.05$ for both), consistent with the bifurcated direction of gene expression change implied by RNA velocity (Figure 4C,E).

TR cells also had lower activity of the pan-photoreceptor regulons LHX3, OTX2, CRX, and NEUROD1 (Figure S5D), consistent with their immature nature.

Overall, high resolution clustering resolved two post-mitotic PR precursor populations with iPRPs having more cone-related and TRs having more rod-related gene expression and regulon activities, consistent with RPC-localized TR cells proceeding towards the ER cluster and iPRP cells continuing towards the larger iPRP population. To our knowledge, this is the first indication of two immediately post-mitotic photoreceptor precursor populations with cone versus rod-biased gene expression (*15*, *16*, *19*).

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

In addition to the immediately post-mitotic iPRP cells, the iPRP cluster included cells bridging the UMAP region between early maturing cones and early maturing rods (Figure 5A). Cells in this region expressed cone markers (*GNAT2*, *THRB*), rod markers (*GNAT1*, *NR2E3*), or, in many cases, both (Figure 5B), resembling a proposed hybrid cone/rod precursor (*4*). To assess if the intermixed cone and rod gene expression reflected 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 iPRP bridge and adjacent ER regions, cells with NRL and THRB regulon signals intermixed, and some cells showed strong signals for both (Figure 5C, arrows). However, bridge region iPRP cells had low NRL regulon z-scores similar to L/M cones and had THRB z-scores above that of adjacent ER cells (Figure 5D). The mean difference between regulon signals increased in rod and cone clusters outside the bridge as NRL or THRB regulons became more dominant (Figure 5D). The iPRP bridge region cells' similar THRB and NRL regulon activities and RNA velocity vectors are consistent either with distinct early cone and rod precursors or with a common cone-rod precursor (*4*).

To determine if cells with early cone and rod RNA co-expression can be detected in the developing retina, we performed multiplex RNA FISH for early cone marker *GNAT2* and rod

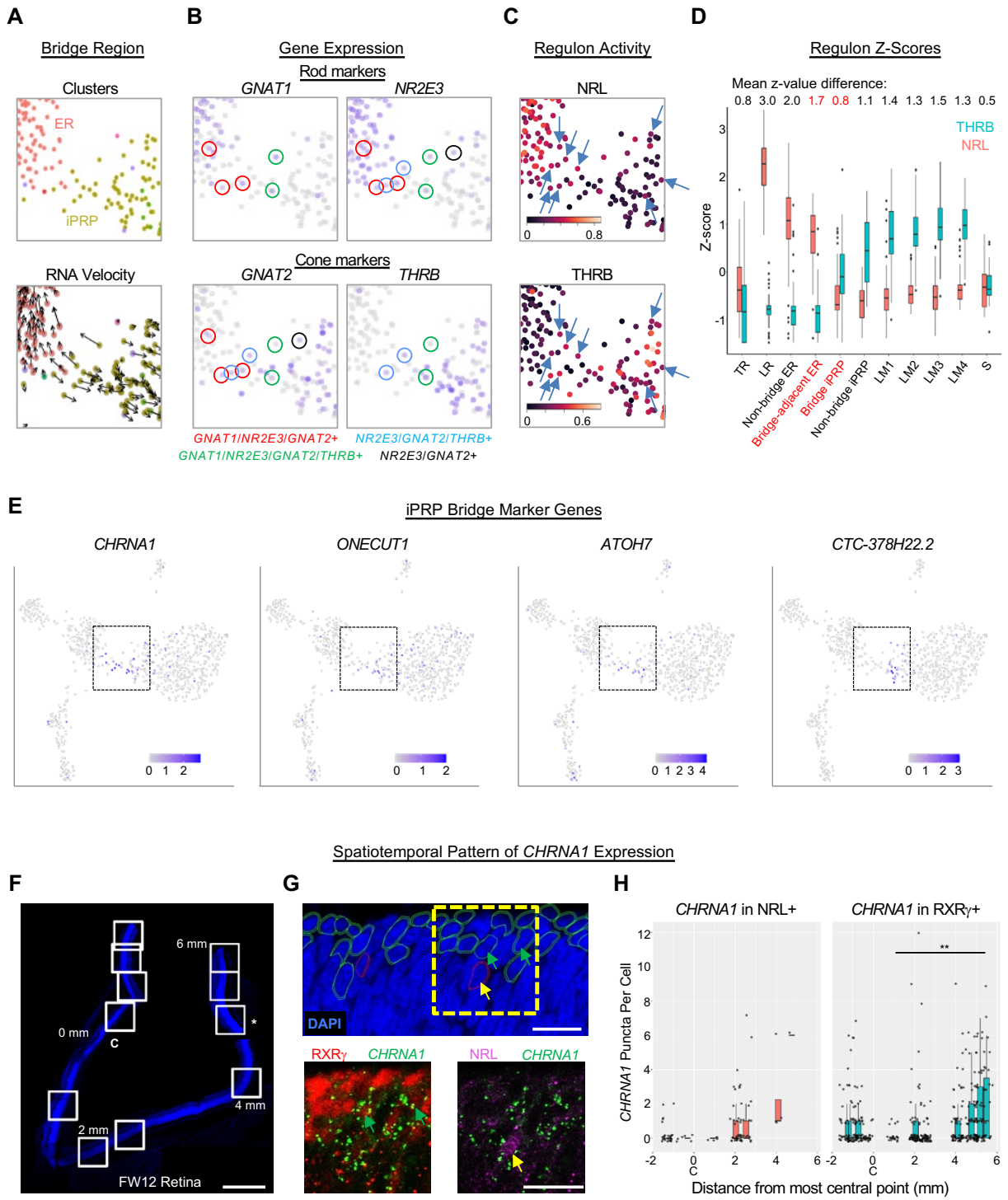

A
Bridge Region
Clusters
ER
iPRP
RNA Velocity
B
Gene Expression
Rod markers
GNAT1
NR2E3
Cone markers
GNAT2
THRB
GNAT1/NR2E3/GNAT2+
GNAT1/NR2E3/GNAT2/THRB+
NR2E3/GNAT2/THRB+
NR2E3/GNAT2+
C
Regulon Activity
NRL
0
0.8
THRB
0
0.6
D
Regulon Z-Scores
Mean z-value difference:
0.8 3.0 2.0 1.7 0.8 1.1 1.4 1.3 1.5 1.3 0.5
THRB
NRL
Z-score
TR
LR
Non-bridge ER
Bridge-adjacent ER
Bridge iPRP
Non-bridge iPRP
LM1
LM2
LM3
LM4
S
E
iPRP Bridge Marker Genes
CHRNA1
ONECUT1
ATOH7
CTC-378H22.2
0 1 2
0 1 2
0 1 2 3 4
0 1 2 3
Spatiotemporal Pattern of CHRNA1 Expression
F
6 mm
0 mm
C
*
4 mm
2 mm
FW12 Retina
G
DAPI
RXRγ
CHRNA1
NRL
CHRNA1
H
CHRNA1 in NRL+
CHRNA1 in RXRγ+
CHRNA1 Puncta Per Cell
**
Distance from most central point (mm)

Figure 5: **An iPRP bridge population with properties of a common cone-rod precursor. A-C.** UMAP bridge region plots displaying high resolution ER and iPRP clusters and RNA velocity (A), rod and cone marker gene expression (B), and NRL and THRB regulon activities (C). In B, circle colors indicate different patterns of GNAT1/NR2E3/GNAT2/THRB co-expression. In C, arrows indicate cells with both regulon signals. **D**. Box plot of Z-score-normalized AUC for each cluster. Bridge region iPRP and ER cells are segregated from other iPRP and ER cells (red labels). Mean of absolute Z-score differences between THRB and NRL regulons displayed at top. **E.** UMAP plots of iPRP bridge region marker genes. **F-H.** Combined RXRγ/NRL immunohistochemical staining and *CHRNA1* FISH of FW12 retina. **F**. 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 in G. Scale bar = 500 µm. **G.** *Top:* Retinal nuclear and cellular segmentation and identification of cells as RXRγ+ (green), NRL+ (red). Yellow box: Field shown below. *Bottom:* RXRγ or NRL immunofluorescence with *CHRNA1* FISH. Arrows: RXRγ+*CHRNA1*+ (green), NRL+*CHRNA1*+ (yellow). Scale bars = 15 µm. **H**. 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).

marker *NR2E3* together with immunostaining for RXRγ, which has high expression in outer neuroblastic layer (NBL) cone precursors and low expression in inner NBL rod precursors (Figure S3E), and for NR2E3, which is detected solely in rods. We examined *GNAT2* and *NR2E3* RNA co-expression in outer NBL RXRγ+ cones and in inner NBL RXRγ^{lo} rods across 13 peripheral fetal retina regions (Figure S6A-C). The analyses revealed that most of the far peripheral (hence, nascent) outer NBL RXRγ+ cones (Figure S6, regions 2 and 13) lacked detectable *GNAT2* and *NR2E3* RNA, whereas those in the more mature central retina (regions 3-12) were uniformly *GNAT2*+. Starting with regions 3 and 11, some *GNAT2*+ cones were also *NR2E3*+ (Figure S6D-E). The proportion of *GNAT2*+, *NR2E3*+ outer NBL cells increased in the more central retina, yet all of these cells had strong RXRγ staining and none had detectable NR2E3 protein, consistent with their having a cone identity (Figure S6D,E). In contrast, most of the inner NBL RXRγ^{lo} cells had either low *NR2E3* RNA without NR2E3 protein, in nascent rods, or prominent *NR2E3* RNA with NR2E3 protein, in maturing rods. However, we did not detect *GNAT2* puncta in inner NBL RXRγ^{lo}, *NR2E3*+ cells, including the most peripheral RXRγ^{lo} cells with low-level *NR2E3* RNA (Figure S6D,F). In summary, most, if not all, photoreceptor precursors with *NR2E3* and *GNAT2* RNA co-expression had high RXRγ and outer NBL positions expected of cone precursors.

To further assess whether iPRP bridge region cells comprise a distinct subpopulation, we sought to identify RNA markers of this region. Among four iPRP marker genes (Figure S5A), *CHRNA1* was mainly expressed in the bridge region and adjacent rods and cones, whereas *ONECUT1* was biased to cone-directed iPRPs, lncRNA *CTC-378H22.2* was expressed in a narrow zone of *THRB*+, *GNAT2*+ cone precursors, and *S100A6* was more widely expressed (Figure 5E, S7A). Bridge region 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 cone-directed cells whereas *DLL3* was broadly expressed, similar to *S100A6* (Figure 5E, S7A).

As *CHRNA1* was the most specific marker of bridge region cone and rod precursors, we evaluated whether *CHRNA1* RNA marked specific cone and rod populations during fetal retinal development using fluorescent *in situ* hybridization combined with RXRγ and NRL immunofluorescent staining. In a FW12 retina, we detected highest *CHRNA1* in the earliest, most peripheral NRL+ rods and RXRγ+ cones and fewer *CHRNA1*+ cells in more mature, central regions ($p < 0.005$ in cones) (Figure 5F-H). The detection of *CHRNA1* RNA in the youngest rods and cones by *in situ* hybridization is consistent with its expression in photoreceptor precursor bridge region cells in scRNA-seq analysis.

To identify factors that regulate transcriptomic states during iPRP cell fate determination, we examined the distribution of the most iPRP-specific regulons identified after high-resolution clustering, OLIG2 and LHX9 (Figure S7B, Table S6). As Olig2 was previously detected in mouse RPCs in which Onecut1 enabled *Thrb* expression and cone fate determination (*41*, *45*), the OLIG2 regulon was expected to be most active in RPCs and in early, immediately post-mitotic iPRPs. While the OLIG2 and LHX9 regulons were indeed active in RPCs, both were also active in a narrow zone of *ONECUT1*+ cone-directed iPRP cells distal to the bridge region and immediately prior to THRB regulon upregulation. These data indicate that the OLIG2 and LHX9 regulons are active in and potentially relevant to human cone fate determination in post-mitotic photoreceptor precursors.

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

After fate commitment, L/M cones initiate a maturation process with outer segment morphogenesis and upregulation of RNAs and proteins related to phototransduction (*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 6A,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 across maturing L/M cones by pseudotemporally ordering the iPRP and L/M cones (Figure 6C). This identified 967 pseudotime-correlated genes (q-value < 0.05, expression > 0.05 in >5% of ordered cells) (Table S6) composing seven gene modules (Figure S7A, B). Among the top 20 pseudotemporally-correlated genes in each module, we identified four lncRNAs that were sequentially expressed across iPRPs and maturing L/M cones (Figure 6D), whereas similar sequential expression was not observed with pseudotime-correlated protein-coding mRNAs. 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 6E,F). Quantitation of FISH puncta defined four cone maturation zones based on expression peaks and significant count differences for each lncRNA: the most peripheral cones with high *CTC-378H22.2* and *HOTAIRM1*, peripheral cones with high *HOTAIRM1* only, cones where all four lncRNAs show low expression, and parafoveal/foveal cones with upregulated *RP13143G15.4*, C*TD-2034I21.2*, and *CTC-378H22.2* (Figure 6G). The detection of foveal *CTC-378H22* by ISH but not in late LM transcriptomes may relate to a lack of the most mature cells in our scRNA-seq analyses. Nevertheless, these data support the concept that specific lncRNA combinations mark maturing L/M cone stages.

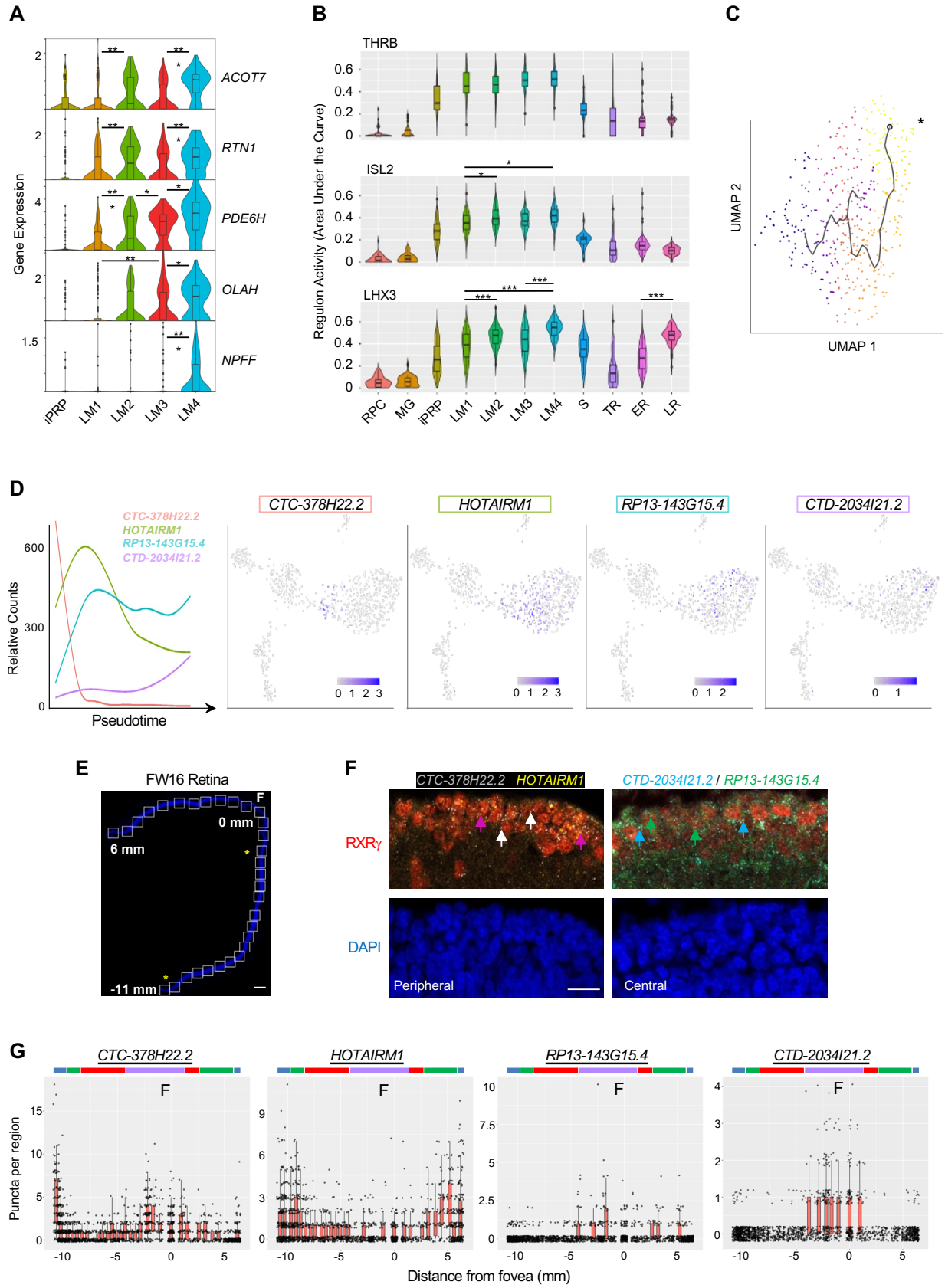

A
Gene Expression
ACOT7
RTN1
PDE6H
OLAH
NPFF
iPRP
LM1
LM2
LM3
LM4
B
Regulon Activity (Area Under the Curve)
THRB
ISL2
LHX3
RPC
MG
iPRP
LM1
LM2
LM3
LM4
S
TR
ER
LR
C
UMAP 2
UMAP 1
D
CTC-378H22.2
HOTAIRM1
RP13-143G15.4
CTD-2034I21.2
Relative Counts
Pseudotime
E
FW16 Retina
0 mm
6 mm
-11 mm
F
CTC-378H22.2 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 6: 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. **D.** Trendlines of relative count expression (*left*) and UMAP plots for lncRNAs correlating with early or late-upregulating modules. Line color matched to label boxes. **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).

Cone intrinsic MYCN and SYK contributions to the proliferative response to pRB loss

We next assessed whether early maturing cone transcriptomic features are conducive to their 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 7A, Table S7). Among cone-enriched genes, the top three enriched ontologies related to translation initiation, protein localization to membrane, and MYC targets (Figure 7B). Although *MYC* RNA was not detected, the upregulation of MYC target genes was of interest given the cone precursors' high MYCN protein expression (Figure 7C) and the critical role of MYCN in the cone precursor proliferative response to pRB loss (*8*, *10*, *46*). Concordantly, the LM cluster had increased *MYCN* RNA ($\log_2FC = 0.54$) and MYCN regulon activity, representing the seventh highest LM cluster regulon specificity score (Figure 7D,E, Table S4).

Among other differentially expressed genes, we noted cone-specific upregulation of *SYK* (Figure 7A,F, Table S7), 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 (*47*), its expression was not previously reported in developing fetal retina. Indeed, *SYK* RNA expression increased at the iPRP stage and into clusters LM3 and 4 and was abolished during the transition from early-maturing to late-maturing L/M cones, in contrast to its minimal expression in the corresponding rod TR, ER, and LR stages (Figure 7G). 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 7H). 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 (*47*) and implies that SYK is a defining feature of the human early cone precursor state. While *MYCN* was also preferentially

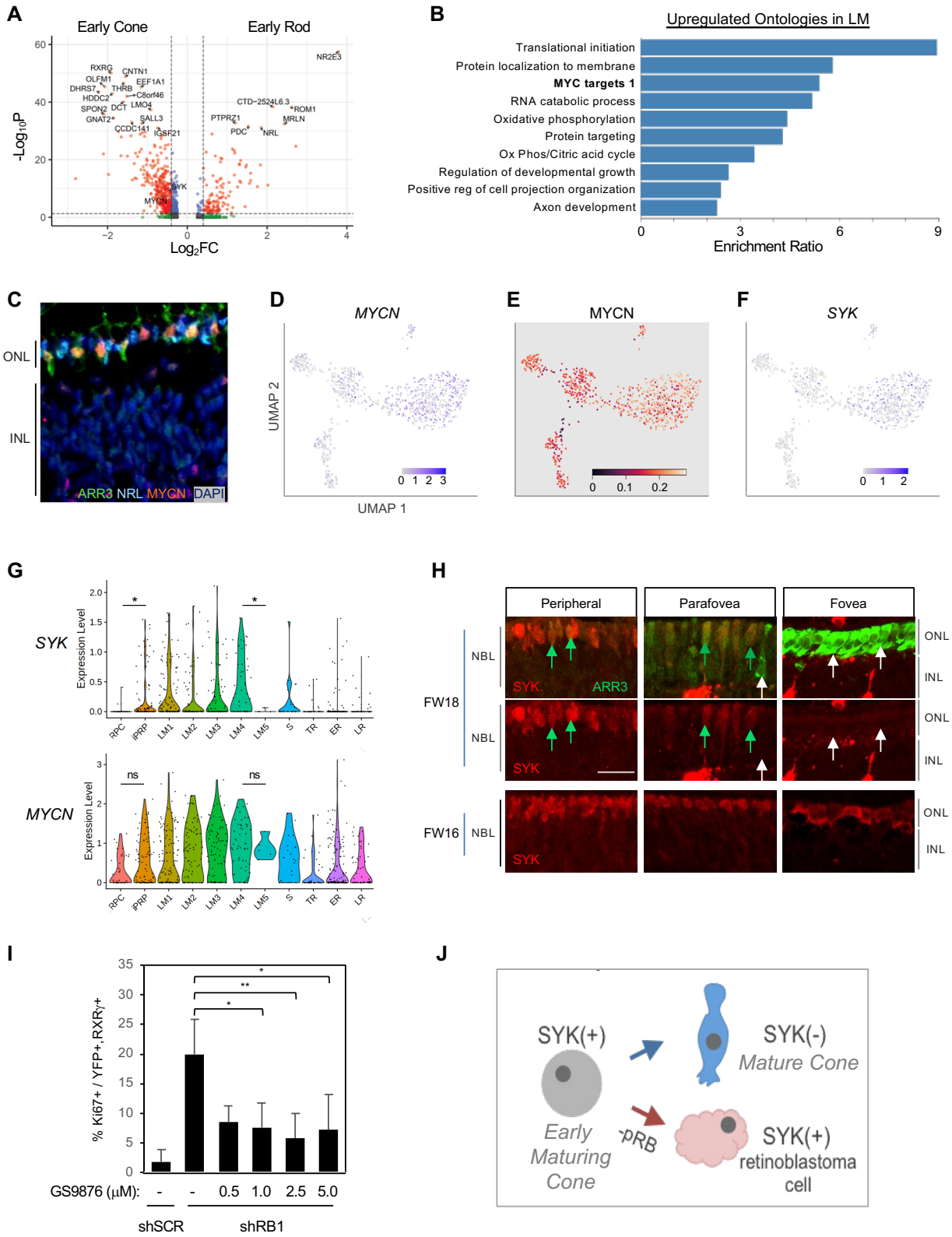

A
Early Cone
Early Rod
-Log10P
Log2FC
RXRG
CNTN1
OLFM1
EEF1A1
DHRS7
THRB
HDDC2
C8orf46
SPON2
DCT
LMO4
GNAT2
SALL3
CCDC141
ISSF21
SYK
MYCN
NR2E3
CTD-2524L6.3
ROM1
PTPRZ1
MRLN
PDC
NRL
B
Upregulated Ontologies in LM
Translational initiation
Protein localization to membrane
MYC targets 1
RNA catabolic process
Oxidative phosphorylation
Protein targeting
Ox Phos/Citric acid cycle
Regulation of developmental growth
Positive reg of cell projection organization
Axon development
Enrichment Ratio
C
ONL
INL
ARR3 NRL MYCN DAPI
D
MYCN
UMAP 2
UMAP 1
E
MYCN
F
SYK
G
SYK
MYCN
Expression Level
RPC
iPRP
LM1
LM2
LM3
LM4
LM5
S
TR
ER
LR
ns
H
Peripheral
Parafovea
Fovea
FW18
FW16
NBL
ONL
INL
SYK
ARR3
I
% Ki67+ / YFP+, RXRγ+
GS9876 (μM):
shSCR
shRB1
J
SYK(+)
Early Maturing Cone
SYK(-)
Mature Cone
pRB
SYK(+)
retinoblastoma cell

Figure 7: Cone intrinsic MYCN and SYK contribute to the proliferative response to pRB loss. A. Volcano plot of differential expression between early rod cluster ER and cone LM, excluding late-maturing OPN1LW+ 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|. **C.** Immunofluorescent staining of MYCN in ARR3+ cones but not NRL+ rods in FW18 retina. **D-F.** UMAP plots of *MYCN* expression (D), MYCN regulon activity (E), and *SYK* expression (F). **G.** *SYK* and *MYCN* gene expression violin plots by cluster. *, p <0.05; ns = not significant (t-test). **H.** Immunohistochemical staining analysis of SYK expression in FW18 and FW16 retinae with rabbit polyclonal antibody (LUMIF-hCAR) for human cone arrestin (ARR3). Green arrow: ARR3+, SYK+. White arrow: ARR3+, SYK-. Scale bar = 25 µm. **H.** Effect of SYK inhibitor GS-9876 on KI67+ expression in shRB1- and YFP transduced RXRγ+ cone cells from dissociated FW16.5 fetal retina. Values represent means of three immunocytochemical 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. **I.** Model of SYK expression and function in cone maturation and retinoblastoma development.

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 7G).

To determine if SYK contributes 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 highly specific SYK inhibitor GS-9876 (*48*) for 12 days, and examined for cone precursor cell cycle entry by co-immunofluorescence 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 7I), indicating that cone precursor intrinsic SYK activity is critical to the proliferative response to pRB loss and is retained in retinoblastoma cells (Figure 7J).

Discussion

This study evaluated transcriptome state changes associated with human cone and rod photoreceptor development using deep full-length scRNA-seq. Whereas prior scRNA-seq of the developing retina provided insights into RPC fate determination and transcriptomic trajectories using 3' end counting (*14–18*, *25*), our experiments enabled deep full-length sequencing for more precise resolution of cell states, identification of cell type-specific regulon activities and transcript isoforms, and detection of previously unrecognized post-mitotic photoreceptor precursor trajectories. Additionally, our cell enrichment strategy provided insight into a cancer-predisposing feature 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 in individual cells. Here, we show that the rod determinant *NRL*, the L/M-cone determinant *THRB*, and the cone maker *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* transcripts that are predicted to encode a truncated NRL (Tr-NRL) lacking a transactivation domain as well as *NRL* isoforms with unusual intra-exon transcription initiation and alternative splicing. As with DD10 (*34*), the Tr-NRL protein 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 specific contexts, not examined in our assays, 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 NRL protein. Thus, the L/M cone precursors' more extensive Tr-NRL expression and *NRL* exon 2 disruptions reveal a multipronged approach to suppress NRL function while downregulating overall *NRL* RNA by less than 4-fold. Similarly, we uncovered cell type-specific differences in *THRB* transcript isoforms, albeit generated through premature transcription termination and 3' UTR expression in late rod versus L/M cone precursor populations. The robust premature transcription termination of *THRB* is consistent with the widespread use of this mechanism to govern expression of diverse transcriptional regulators (*49*). A limitation of these studies is that individual cells may express a small and varying spectrum of transcript isoforms at any one time; while this was mitigated by combining cDNAs of cells deemed to be in similar states based on short-read sequencing, analysis of more cells would 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 (*29*). Our deep sequencing approach uncovered discrete cell states distinguished by gene expression and, more notably, regulon activities. For example, early and late maturing rod populations were distinguished by the latter's increased expression of the NRL, ATF4, CRX, and LHX3

regulons and decreased HMX1 and RAX regulons. In contrast, 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*, *50*), and a late maturing cone group with upregulated L/M opsin gene expression and decreased RAX activity (Figure 1D,E). 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 both the cone opsin and rhodopsin promoters (*51*, *52*).

Increased clustering resolution further divided the L/M cone cluster into four subgroups that had only subtle maturation-associated changes in marker gene expression and regulon activity, consistent with LM1-4 comprising different stages of a graded maturation process. However, pseudotemporal analysis identified a maturation trajectory with successive expression of lncRNAs that was validated in developing retinal tissue. *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 although without sequential expression (*15*, *17*, *53*). 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 the immature photoreceptor precursor (iPRP) and transitional rod (TR) populations. iPRP and TR cells positioned adjacent to RPC/MG cells in UMAP plots showed higher cone-like or rod-like gene expression and regulon activities, respectively, and had RNA velocities directed toward distinct cone and rod populations (Figure 4C-E). Prior droplet-based scRNA-seq analyses have not to our knowledge distinguished immediately post-mitotic cone- and rod-directed photoreceptor precursors. However, scRNA-seq and scATAC-seq analyses showed distinct early neurogenic precursors directed in part towards cones and late neurogenic precursors directed in part towards rods, which may comprise similar populations (*14*, *15*, *19*). As one possibility, early iPRPs may comprise a cone

directed subset of early neurogenic cells while early TRs comprise a rod-directed subset of late neurogenic cells.

Although not identified as a distinct cluster, our analyses also revealed iPRP 'bridge region' cells that co-expressed rod and cone genes and regulons (Figure 3F). This population resembled the T3 transition and early neurogenic precursor states identified via 3' end-counting (*15*, *16*), with expression of *ATOH7*, *DLL3*, and the new marker *CHRNA1*. Bridge region cells with velocity directed towards early cones expressed *THRB* and *ONECUT1* RNAs and had OLIG2 regulon activity; as these elements were proposed to promote cone fate in lineage-restricted RPCs (*41*), their expression in the iPRP population suggests a similar role in post-mitotic photoreceptor precursors. Conversely, the status of bridge region cells with velocity directed towards early rods is less clear and could reflect TR-derived cells with transcriptomic properties similar to cone-directed iPRPs or iPRP-derived cells that follow an alternative route to rods. Our finding that cells co-expressing cone marker *GNAT2* and rod marker *NR2E3* RNAs had outer NBL locations and high RXRγ supports the notion that such cells are cone-committed.

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 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 contributes to cone precursor proliferation and retinoblastoma tumorigenesis. We further found that cone precursors highly express *SYK* RNA and protein. SYK was previously detected in human retinoblastoma and in *RB1*-null retinal organoid tumor-like masses but not in healthy tumor-associated retina (*47*, *54*). The high SYK expression preceding early cone precursor maturation and its decline to undetectable levels coinciding with late maturation implies that it is a retinoblastoma cell-of-origin feature that is lost in later cone development. SYK inhibition impaired pRB-depleted cone precursor cell cycle

entry, suggesting that native SYK expression, rather than *de novo* induction, contributes to the cone precursors' initial proliferation (Figure 7J).

Our study has distinct limitations, including the relatively low number of cells examined as compared to droplet-based studies. However, this is mitigated in part by deep sequencing, with >8,000 genes detected per cell, which may be used to clarify the transcriptomes of rare cell populations (*55*). Long-read scRNA-seq of larger cell numbers along with lineage tracing in model systems is needed to further resolve human photoreceptor developmental trajectories.

In summary, through deep, full-length RNA sequencing, we identified photoreceptor cell type-specific differences in transcript isoform expression and immediately post-mitotic photoreceptor precursor states with distinctive gene expression and regulon activities. The discrimination of these states suggests novel developmental trajectories and cone precursor features that underlie retinoblastoma predisposition.

Materials and Methods

Primary human tissue

Human fetal samples were provided by Advanced Bioscience Resources, Inc. or collected under Institutional Review Board approval at the University of Southern California and Children's Hospital Los Angeles. 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 (*56*). Each retina subjected to sequencing was from a different donor.

Retinal dissociation and RNA-sequencing

Single cell isolation

Retinae were dissected and dissociated in ice-cold phosphate-buffered saline (PBS) 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 (IMDM; Corning), 10% FBS, 0.28 U/mL insulin, 55 µM β-mercaptoethanol (Sigma-Aldrich), glutamine, penicillin, and streptomycin] 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 HBSS, tubes were recombined and cells counted.

Fluorescence-activated cell sorting (FACS) isolation of single cells and cDNA synthesis

Single cells were FACS-isolated as described (*10*) with differences as follows. After removing cells for an unstained control, dissociated cells were centrifuged and resuspended in a 5% fetal bovine serum in PBS (FBS/PBS) with CD133-PE (Miltenyi Biotec 130-080-801), 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 hr, 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 1.2 µl lysis buffer droplets on parafilm-covered glass slides (Lee et al. *In preparation*). Gating was set to remove low forward- and side-scatter cells before collecting a CD133-high, CD44/49b-low population. Upon collection of eight cells, droplets were transferred on ice into individual low-retention PCR tubes (Bioplastics K69901, B57801). Sample tubes were reverse transcribed 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 (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 immunologically stained and FACS isolated as above and collected into a single 1.5 ml tube. During collection, the microfluidics chip was primed per Clontech kit instructions for the C1 system. Cells were centrifuged and resuspended at 400-800 cells/µl, combined with C1 suspension reagent, and loaded onto the C1 chip. After flow through, the chip was imaged at each site to identify cell number in each well and loaded for sample preparation. Sequencing was performed 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 fresh low-retention tube strips and stored at -20°C until quantitation and library preparation.

Quality control, library preparation and sequencing

DNA quantitation was performed using Quant-iT PicoGreen dsDNA assay and a subset of samples were checked on the Bioanalyzer platform for quality and to determine PicoGreen thresholds for inclusion in libraries (typically 0.05 ng/µl). Libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1096). The Seq 1 retinal cell libraries were sequenced on the Illumina NextSeq 500 (2x75) and all others on Illumina HiSeq 4000 (2x75).

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 individually re-amplified using the 5' PCR Primer II A with CloneAmp HiFi PCR Premix (Clontech Laboratories, Inc. A Takara Bio Company, #639298) 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 used for separate RACE PCR reactions with gene-specific *NRL, RXRγ, TRβ1 and TRβ2* primers (see below) 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. 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. Forward and reverse barcode sequences are provided by ONT. 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 flowcell (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.

Primers used for RACE reactions:

5' PCR Primer II A: 5'-AAGCAGTGGTATCAACGCAGAGT-3'

5'RACE: 5'-AAGCAGTGGTATCAACGCAGAGTACGGG -3'

Rev NRL-Full ex3: 5'- GGTTTAGCTCCCGCACAGACATCGAGAC -3'

Rev RXR ex5: 5'- GAAGAACCCTTTGCAGCCTTCACAACTG -3'

3' RACE: 5'- AAGCAGTGGTATCAACGCAGAGTACTTTT-3'

Forw TRb1 ex4: 5'- GCCTTACAGCCTGGGACAAACC-3'

Forw TRb2 ex1: 5'-CCCTGGAAACATGTTTAAAAGCAAGGACT-3'

Cells analyzed in RACE reactions:

Age of retinae	**FW13**	**FW15**	**FW16**	**FW18**	**FW19**	**Total**
# cells from ER	1	7	0	6	9	23
# cells from LM	1	7	0	2	11	21
# cells from LR	0	0	2	2	1	5

Computational Methods

Read processing and dimensionality reduction

Illumina short-read sequencing FASTQ files had adapter sequences removed using the *trimgalore* wrapper for *Cutadapt* (*57*). Trimmed FASTQ files were used as input for *HISAT2* (*58*) 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* (*59*) and *Tximport* (*60*), 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 (*61*) workflow accessible at https://github.com/whtns/ARMOR.

For short-read sequencing, dimensionality reduction and data visualization were performed using the Seurat (V3) toolset (*62*, *63*). To exclude technical effects between distinct tissue isolation and sequencing batches, expression counts from all 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 uniform manifold approximation and projection (UMAP) embeddings (*64*, *65*). 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) (*66*).

Read coverage plots for genes of interest were generated using *wiggleplotr* (*67*) to visualize BigWig files with ENSEMBL isoforms. Individual exon counts were identified using *DEXseq* (*68*), 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.

Analysis of ONT long-read sequencing data was performed with the nf-core nanoseq version 3.1.0 pipeline (*69*) (https://nf-co.re/nanoseq/3.1.0). Quality control on raw reads was conducted using FastQC. Reads were aligned using minimap2 (*70*). 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 (*71*). 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 analysis

To identify marker features for each cluster, Wilcoxon-rank sum tests and specificity scores were computed using the *genesorteR* (*72*) R package. Only genes with an adjusted p-value (pAdj) <0.05 were considered for further analyses and the top five markers for each cluster were displayed where possible. Differential expression was performed between groups of interest

using the Seurat FindMarkers function with the Wilcoxon-rank sum test method on log-normalized counts. Adjusted p-values (Bonferroni correction on all features) of <0.05 were considered significant. Results were displayed as volcano plots made with *EnhancedVolcano* (*73*).

Overrepresentation analyses were performed with WebGestalt (*74*). 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) (*29*, *75*). 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* (*37*). 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 (*76*). The SeuratWrappers package provided integration between Seurat's final integrated dimensional reduction and Monocle. The data was subset to include only cone-directed iPRP 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, then 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, 2-3 sections per slide 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, then the membrane was cut from its frame and transferred into a mold containing 1:2 OCT:30% sucrose and frozen.

Immunostaining

For SYK and ARR3 co-immunohistological 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 antibodies, mouse anti-SYK and rabbit anti-ARR3 (LUMIF-hCAR), 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 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 washing steps.

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). Cryostats were cleaned with a mixture of 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 Resource tables, 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 *NRL* FISH, samples were washed with DAPI in the third and final wash, mounted and then imaged as above. After this first imaging, coverslip was removed, slides were washed in PBS, and then were 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 (*77*) was used to identify nuclei, predict cell expansion, classify cells, and count FISH puncta within those regions. Apical retina regions containing RXRγ+ and/or NRL+ cells were visually selected for evaluation within each image. The StarDist extension (*78*) 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 manually reviewed and poorly identified cells (*e.g.*, cells with partial, multiple, or overlapping nuclei) were 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 by visual assessment. 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. For the inner layer, the number of *NR2E3* RNA positive cells and NR2E3 protein positive cells were counted.

Lentivirus production

Four plasmids were each used in the production of three different lentiviral constructs (pMDL, pVSVG, pREV and a lentiviral vector) (*79*). HEK293T cells were cultured in 15 cm dishes in 293T media [Dulbecco's Modified Eagle's Medium (DMEM), 10% FBS, 1% penicillin-streptomycin] and split to the desired plate number the night before transfection with no antibiotics. Plasmids were mixed at 10 µg pMDL, 5 µg pREV, 5 µg pVSVG, and 20 µg vector in 3 ml of fresh DMEM without serum for each plate. Separately, 160 µl of 0.6 µg/µl polyethyleneimine (PEI) and 3 ml DMEM were mixed per plate and allowed to stand for 5 min. Both mixtures were combined and allowed to incubate at RT for at least 30 min. All 293T plates were treated with 2 ml trypsin (Corning 25-052-CI) to single cell dissociation, diluted and harvested in 3 ml DMEM + 10% FBS, counted, and the volume adjusted if needed to contain approximately 30 million cells/plate. The cell mixture and PEI/plasmid mix were combined and mixed for several minutes before being evenly distributed to the original number of 15 cm plates. DMEM + 10% FBS was added to bring the final volume to 25 mL. After 16-24 h at 37°C, media was exchanged for 20 mL of UltraCULTURE medium (Lonza Inc.) containing 1% HEPES, 1% GlutaMax and 1% Penicillin-Streptomycin. 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).

NRL isoform analyses

Proteasome inhibition of CHLA-VC-RB-31

10^6 CHLA-VC-RB-31 cells (*35*) 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. Proteasome inhibitor MG132 (Sigma-Aldrich 474790) was added at 23 h to a final concentration of 10 µM and cells incubated for an additional 4 h, collected and centrifuged at 300 x g for 4 min, resuspended in 60 µl cold 1x RIPA buffer (10x RIPA (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 before collecting supernatant. 15 µl of cell lysate 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) at 200 v for 50 min 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, and incubated overnight in 0.1% dry milk TBS-T solution containing primary antibody at 4°C with slow agitation on a rocker, followed by three 15 min TBS-T washes, HRP-conjugated secondary antibody incubation for 1 hr at RT in the same 0.1% milk solution, washing membrane six times for 15 min, incubating in chemiluminescent substrate (Thermo Fisher, 34094) and imaging.

NRL isoform and reporter constructs

The pcDNA4-His-Max-C-Nrl (*80*) (gift of A. Swaroop) CMV promoter was replaced with EF1α promoter to produce pcDNA4-His-Max-C-EF1a-FL-NRL. Using In-Fusion (Takara Bio) the His-Max tag was removed to produce full-length NRL expression plasmid pCDNA4-C-EF1a-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-EF1a-Tr-NRL (primers #1,2,3,4). pcDNA-C-EF1a-empty transfection carrier plasmid was made by removing NRL coding sequence from pCDNA4-C-EF1a-Fl-NRL (primers #5,6). pGL3-PDE6B-146-luc was produced by amplifying a 146 bp promoter region of *PDE6B* (*81*) 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-FL-NRL was produced by amplification of the FL-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 Resources table.

NRL expression construct and antibody validation

pcDNA4-C-EF1α-NRL constructs were transfected into HEK293T cells. 4 µg plasmid in 1 ml DMEM was combined with 1 ml DMEM + 12 µg PEI 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, then after two days samples were collected, lysed, and 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).

Concentrated pUltra-EGFP-NRL-205 and control pUltra-EGFP lentivirus were adjusted to equal virus concentrations in lentiviral collection media, diluted 1:4 in 293T media with 10% FBS and Polybrene (5 µg/ml), and 1 ml incubated with 250,000 HEK293T cells per well in a 12-well plate. After overnight infection, media was replaced with fresh 293T media, and after two days samples were attached to poly-L-Lysine-coated glass coverslips for 2 h, fixed in 4% PFA for 10 min, washed in 1x PBS three times, and stored at 4°C before immunostaining.

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 pcDNA4C-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

As previously 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 as above 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, then 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, then cells were thresholded for GFP, RXRγ, and NRL.

RB1 knockdown and SYK inhibitor treatment

FW18.5 retina was cut into peripheral and central regions (outer 1/3 and inner 2/3 of the tissue), partially dissociated with short papain treatment, 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 concentrations in the retina culture medium 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) 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 immunohistological staining.

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.

Resource tables

Antibodies

ANTIBODY	SOURCE	IDENTIFIER
Goat polyclonal anti-NRL (WB: 1:2000-4000)	R&D Systems	1) CAT# AF2945, RRID: AB_2155098
Mouse monoclonal anti-NRL (WB:1:250, IF:1:50)	Santa Cruz	CAT# SC-374277, RRID: AB_10991100
Mouse monoclonal anti-RXRγ (1:200)	Santa Cruz	CAT# SC-514134, RRID: AB_2737293
Mouse monoclonal anti-SYK (1:200)	Santa Cruz	CAT# SC1240, RRID: AB_628308
Rabbit polyclonal anti-ARR3 (1:5000)	Cheryl Craft (Li et al., 2003; Zhang et al., 2001)	LUMI-F - hCAR
Rabbit polyclonal anti-RXRγ (1:800)	Santa Cruz	CAT# SC-555, RRID: AB_2269865
Mouse monoclonal anti-KI67 (1:200)	BD Bioscience	**2)** CAT# 550609, RRID: AB_393778
Goat polyclonal anti-GFP (1:500)	Abcam	**3)** CAT# ab6673, RRID: AB_305643
Mouse monoclonal anti-CD133-PE (1:50)	Miltenyi Biotec	4) CAT# 130-113-108, RRID: AB_2725937
Mouse monoclonal anti-CD44-FITC (1:50)	BD Biosciences	5) CAT# 555478, RRID: AB_395870
Mouse monoclonal anti-CD49b-FITC (1:10)	BD Biosciences	6) CAT# 555498, RRID: AB_395888
Donkey polyclonal anti-goat IgG Alexa Fluor 488 (1:300)	Jackson Laboratories	7) CAT# 705-545-147, RRID: AB_2336933
Donkey polyclonal anti-mouse IgG Alexa Fluor 680 (1:500)	Life Technologies	8) CAT# A10038, RRID: AB_2534014
Donkey polyclonal anti-mouse IgG Alexa Fluor 680 (1:500)	Invitrogen	9) CAT# A11057, RRID: AB_2534104

Tissue Samples

SAMPLE	SOURCE	IDENTIFIER
Human fetal retina	Family Planning Associates, Los Angeles, CA	Not applicable
Human fetal retina	Advanced Bioscience Resources, Alameda, CA	Not applicable

In Situ RNA Probes

PROBE TARGET	SOURCE	IDENTIFIER
FL-NRL exon 1, 2	Molecular Instruments Inc.	Target sequence Ensembl ID: ENST00000619224.1
CHRNA1	Molecular Instruments Inc.	Target sequence NCBI accession number: NM_000079.4
RP13-143G15.4	Molecular Instruments Inc.	
CTC-378H22.2	Molecular Instruments Inc.	Target sequence Ensembl ID: ENST00000559786.1
HOTAIRM1	Molecular Instruments Inc.	Target sequence NCBI accession number: NR_038366.1
CTD-2034I21.2	Molecular Instruments Inc.	Target sequence NCBI accession number: XR_001752169.1
GNAT2	Molecular Instruments Inc.	Target sequence NCBI accession number: NM_005272.5
NR2E3	Molecular Instruments Inc.	Target sequence NCBI accession number: NM_01249.4

Chemicals and Assays

PRODUCT	SOURCE	IDENTIFIER
GS-9876 SYK inhibitor	MedChemExpress	CAT# HY-109091
SMARTer Ultra Low RNA Kit for the Fluidigm C1 System SMART-Seq V4 Ultra Low Input RNA Kit	Clontech (Takara Bio) Takara Bio	CAT# 634835/634935 CAT# 63891
Nextera XT DNA Library Preparation Kit	Illumina	CAT# FC-131-1096
Nano-Glo Dual-Luciferase Reporter Assay System	Promega, Inc	CAT# N1610

Cell Lines

CELL LINE	SOURCE	IDENTIFIER

CHLA-VC-RB-31	Stachelek et al., 2022	N/A
HEK-293T	ATCC	N/A
NIH-3T3	ATCC	N/A

Oligonucleotides

PRIMER NAME	SOURCE	SEQUENCE
10) Del-His F	N/A	CCGAAACCATGGCCCTGCCCCCCAGC
11) Del-His R	N/A	GGGCCATGGTTTCGGAGGCCGTCCG
12) NRL-no-His F	N/A	CCGAAACCATGTCTGTGCGGGAGCTAAACC
13) NRL-no-His R	N/A	CAGACATGGTTTCGGAGGCCGTCCG
14) pcDNA del NRL F	N/A	CCGAAACCGCCGTTCAGAGCACCTTGTGG
15) pcDNA del NRL R	N/A	GAACGGCGGTTTCGGAGGCCGTCCG
16) PDE -93 F IF	N/A	TCTTACGCGTGCTAGAGCGCAGGCCCCCATTTG
17) PDE +53 R IF	N/A	CTTAGATCGCAGATCGGTGGCTGCCTGTCCCTG
18) pGL3-sv40 SDM F	N/A	CTGCGATCAAGCTTGGCATTCCGGTACTG
19) pGL3-sv40 SDM R	N/A	CAAGCTTGATCGCAGATCGGTGGCTG
20) pGL3 del SV40 R	N/A	CAAGCTTGATCGCAGATCTCGAGCCC
21) FL-NRL IF pU-G F	N/A	GCCTTCTAGAGGATCCATGTCTGTGCGGGAGCTAAACC
22) FL-NRL IF pU-G R	N/A	CGCCGGAGCCGGATCCTCAGAGGAAGAGGTGGGAGGG

Recombinant DNA

PLASMID NAME	SOURCE	IDENTIFIER
pLKO.1C-YFP	Lee and Cobrinik, 2020	N/A
pLKO.1C-YFP-shRB1-733	Lee and Cobrinik, 2020	N/A
pcDNA4-His-Max-C-Nrl	Gift from A. Swaroop. Cheng et al., 2004	N/A
cDNA4-His-Max-C-EF1a-FL-NRL	This study	N/A
pcDNA4-C-EF1a-FL-NRL	This study	N/A
pcDNA4-C-EF1a-Tr-NRL	This study	N/A
pcDNA4-c-EF1a	This study	N/A
pGL3-SV40	Promega	CAT# E1761, RRID: Addgene_173953
pGL3-PDE6B-146	This study	N/A

pGL3-empty	This study	N/A
pUltra-EGFP-FL-NRL	This study	N/A
pUltra-EGFP	Lou et al., 2012	RRID: Addgene_24129

Software

SOFTWARE	SOURCE	LINK
TrimGalore		https://github.com/FelixKrueger/TrimGalore
Cutadapt	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/
HISAT2	Kim et al., 2019	https://github.com/DaehwanKimLab/hisat2
StringTie	Pertea et al., 2015	minmmhttp://ccb.jhu.edu/software/stringtie/
Snakemake/ARMOR	Orjuela et al., 2019	https://github.com/whtns/ARMOR
Minimap2	Li, 2018	https://lh3.github.io/minimap2/
WebGestalt	Liao et al., 2019	http://www.webgestalt.org/
pySCENIC	Van de Sande et al., 2020	https://github.com/aertslab/pySCENIC
Tximport	Soneson et al., 2015	https://bioconductor.org/packages/release/bioc/html/tximport.html
Seurat (v3)	Stuart et al., 2019	https://satijalab.org/seurat/index.html
EnhancedVolcano	Blighe et al., 2018	https://github.com/kevinblighe/EnhancedVolcano
genesorteR (v0.4.3)	Ibrahim and Kramann, 2019	http://github.com/mahmoudibrahim/genesorteR
Wiggleplotr (v1.13.1)	Alasoo, 2019	https://github.com/kauralasoo/wiggleplotr
DEXSeq	Anders et al., 2012	https://bioconductor.org/packages/release/bioc/html/DEXSeq.html
velocytoR	La Manno et al., 2018	https://velocyto.org/
Monocle3	Cao et al., 2019	https://cole-trapnell-lab.github.io/monocle3/
QuPath	Bankhead et al., 2017	https://qupath.github.io/
StarDist (QuPath implementation)	Schmidt et al., 2018	https://qupath.readthedocs.io/en/stable/docs/advanced/stardist.html

References

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

2. J. A. Brzezinski, T. A. Reh, Photoreceptor cell fate specification in vertebrates. *Development* **142**, 3263–3273 (2015).

3. A. Swaroop, D. Kim, D. Forrest, Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat Rev Neurosci* **11**, 563–576 (2010).

4. L. Ng, A. Lu, A. Swaroop, D. S. Sharlin, A. Swaroop, D. Forrest, Two Transcription Factors Can Direct Three Photoreceptor Outcomes from Rod Precursor Cells in Mouse Retinal Development. *The Journal of Neuroscience* **31**, 11118–11125 (2011).

5. A. Hendrickson, D. Possin, L. Vajzovic, C. A. Toth, Histologic Development of the Human Fovea From Midgestation to Maturity. *Am J Ophthalmol* **154**, 767-778.e2 (2012).

6. A. Hoshino, R. Ratnapriya, M. J. Brooks, V. Chaitankar, M. S. Wilken, C. Zhang, M. R. Starostik, L. Gieser, A. La Torre, M. Nishio, O. Bates, A. Walton, O. Bermingham-McDonogh, I. A. Glass, R. O. L. Wong, A. Swaroop, T. A. Reh, Molecular Anatomy of the Developing Human Retina. *Dev Cell* **43**, 763-779.e4 (2017).

7. M. Georgiou, A. G. Robson, K. Fujinami, T. A. C. de Guimarães, Y. Fujinami-Yokokawa, M. Daich Varela, N. Pontikos, A. Kalitzeos, O. A. Mahroo, A. R. Webster, M. Michaelides, 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 (2024).

8. D. Cobrinik, Retinoblastoma Origins and Destinations. *New England Journal of Medicine* **390**, 1408–1419 (2024).

9. K. A. Hussey, S. E. Hadyniak, R. J. Johnston, Patterning and Development of Photoreceptors in the Human Retina. *Front Cell Dev Biol* **10**, 1–21 (2022).

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

11. H. P. Singh, S. Wang, K. Stachelek, S. Lee, M. W. Reid, M. E. Thornton, C. M. Craft, B. H. Grubbs, D. Cobrinik, Developmental stage-specific proliferation and retinoblastoma genesis in RB-deficient human but not mouse cone precursors. *Proc Natl Acad Sci U S A* **115**, E9391–E9400 (2018).

12. S. Blackshaw, S. Harpavat, J. Trimarchi, L. Cai, H. Huang, W. P. Kuo, G. Weber, K. Lee, R. E. Fraioli, S.-H. Cho, R. Yung, E. Asch, L. Ohno-Machado, W. H. Wong, C. L. Cepko, Genomic Analysis of Mouse Retinal Development. *PLoS Biol* **2**, e247 (2004).

13. C. B. Mellough, R. Bauer, J. Collin, B. Dorgau, D. Zerti, D. W. P. Dolan, C. M. Jones, O. G. Izuogu, M. Yu, D. Hallam, J. S. Steyn, K. White, D. H. Steel, M. Santibanez-Koref, D. J. Elliott, M. S. Jackson, S. Lindsay, S. Grellscheid, M. Lako, An integrated transcriptional analysis of the developing human retina. *Development* **146** (2019).

14. B. S. Clark, G. L. Stein-O'Brien, F. Shiau, G. H. Cannon, E. Davis-Marcisak, T. Sherman, C. P. Santiago, T. V. Hoang, F. Rajaii, R. E. James-Esposito, R. M. Gronostajski, E. J. Fertig, L. A. Goff, S. Blackshaw, 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 (2019).

15. Y. Lu, F. Shiau, W. Yi, S. Lu, Q. Wu, J. D. Pearson, A. Kallman, S. Zhong, T. Hoang, Z. Zuo, F. Zhao, M. Zhang, N. Tsai, Y. Zhuo, S. He, J. Zhang, G. L. Stein-O'Brien, T. D. Sherman, X. Duan, E. J. Fertig, L. A. Goff, D. J. Zack, J. T. Handa, T. Xue, R. Bremner, S. Blackshaw, X. Wang, B. S. Clark, Single-Cell Analysis of Human Retina Identifies Evolutionarily Conserved and Species-Specific Mechanisms Controlling Development. *Dev Cell* **53**, 473-491.e9 (2020).

16. A. Sridhar, A. Hoshino, C. R. Finkbeiner, A. Chitsazan, L. Dai, A. K. Haugan, K. M. Eschenbacher, D. L. Jackson, C. Trapnell, O. Bermingham-McDonogh, I. Glass, T. A. Reh, Single-Cell Transcriptomic Comparison of Human Fetal Retina, hPSC-Derived Retinal Organoids, and Long-Term Retinal Cultures. *Cell Rep* **30**, 1644-1659.e4 (2020).

17. D. F. Buenaventura, A. Corseri, M. M. Emerson, Identification of genes with enriched expression in early developing mouse cone photoreceptors. *Invest Ophthalmol Vis Sci* **60**, 2787–2799 (2019).

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

19. P. Lyu, T. Hoang, C. P. Santiago, E. D. Thomas, A. E. Timms, H. Appel, M. Gimmen, N. Le, L. Jiang, D. W. Kim, S. Chen, D. F. Espinoza, A. E. Telger, K. Weir, B. S. Clark, T. J. Cherry, J. Qian, S. Blackshaw, Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. *Cell Rep* **37**, 109994 (2021).

20. Y. Omori, S. Kubo, T. Kon, M. Furuhashi, H. Narita, T. Kominami, A. Ueno, R. Tsutsumi, T. Chaya, H. Yamamoto, I. Suetake, S. Ueno, H. Koseki, A. Nakagawa, T. Furukawa, 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 (2017).

21. G. Gagliardi, K. Ben M'Barek, A. Chaffiol, A. Slembrouck-Brec, J.-B. Conart, C. Nanteau, O. Rabesandratana, J.-A. Sahel, J. Duebel, G. Orieux, S. Reichman, O. Goureau, Characterization and Transplantation of CD73-Positive Photoreceptors Isolated from Human iPSC-Derived Retinal Organoids. *Stem Cell Reports* **11**, 665–680 (2018).

22. T. W. Chen, G. Chen, L. J. Funkhouser, S.-C. Nam, Membrane domain modulation by Spectrins in Drosophila photoreceptor morphogenesis. *genesis* **47**, 744–750 (2009).

23. C. A. Burger, D. Jiang, R. D. Mackin, M. A. Samuel, Development and maintenance of vision's first synapse. *Dev Biol* **476**, 218–239 (2021).

24. D. Sokolov, E. R. Sechrest, Y. Wang, C. Nevin, J. Du, S. Kolandaivelu, Nuclear NAD+-biosynthetic enzyme NMNAT1 facilitates development and early survival of retinal neurons. *Elife* **10**, 1–26 (2021).

25. S. W. Lukowski, C. Y. Lo, A. A. Sharov, Q. Nguyen, L. Fang, S. S. Hung, L. Zhu, T. Zhang, U. Grünert, T. Nguyen, A. Senabouth, J. S. Jabbari, E. Welby, J. C. Sowden, H. S. Waugh, A. Mackey, G. Pollock, T. D. Lamb, P. Wang, A. W. Hewitt, M. C. Gillies, J. E. Powell, R. C. Wong, A single-cell transcriptome atlas of the adult human retina. *EMBO J* **38**, 1–15 (2019).

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

27. A. Kallman, E. E. Capowski, J. Wang, A. M. Kaushik, A. D. Jansen, K. L. Edwards, L. Chen, C. A. Berlinicke, M. Joseph Phillips, E. A. Pierce, J. Qian, T. H. Wang, D. M. Gamm, D. J. Zack, Investigating cone photoreceptor development using patient-derived NRL null retinal organoids. *Commun Biol* **3**, 82 (2020).

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

29. B. Van de Sande, C. Flerin, K. Davie, M. De Waegeneer, G. Hulselmans, S. Aibar, R. Seurinck, W. Saelens, R. Cannoodt, Q. Rouchon, T. Verbeiren, D. De Maeyer, J. Reumers, Y. Saeys, S. Aerts, A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247–2276 (2020).

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

31. M. R. Roberts, A. Hendrickson, C. R. McGuire, T. A. Reh, 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 (2005).

32. J. S. Friedman, H. Khanna, P. K. Swain, R. DeNicola, H. Cheng, K. P. Mitton, C. H. Weber, D. Hicks, A. Swaroop, 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 (2004).

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

34. A. Rehemtulla, R. Warwar, R. Kumar, X. Ji, D. J. Zack, A. Swaroop, The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proceedings of the National Academy of Sciences* **93**, 191–195 (1996).

35. K. Stachelek, N. Harutyunyan, S. Lee, A. Beck, J. Kim, L. Xu, J. L. Berry, A. Nagiel, C. P. Reynolds, A. L. Murphree, T. C. Lee, J. G. Aparicio, D. Cobrinik, Non-synonymous, synonymous, and non-coding nucleotide variants contribute to recurrently altered biological processes during retinoblastoma progression. *Genes Chromosomes Cancer* **62**, 275–289 (2023).

36. H. Khanna, M. Akimoto, S. Siffroi-Fernandez, J. S. Friedman, D. Hicks, A. Swaroop, Retinoic Acid Regulates the Expression of Photoreceptor Transcription Factor NRL. *Journal of Biological Chemistry* **281**, 27327–27334 (2006).

37. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

38. K. Roesch, A. P. Jadhav, J. M. Trimarchi, M. B. Stadler, B. Roska, B. B. Sun, C. L. Cepko, The transcriptome of retinal Müller glial cells. *Journal of Comparative Neurology* **509**, 225–238 (2008).

39. X. Pereiro, A. M. Miltner, A. La Torre, E. Vecino, Effects of Adult Müller Cells and Their Conditioned Media on the Survival of Stem Cell-Derived Retinal Ganglion Cells. *Cells* **9**, 1759 (2020).

40. J. C. Walcott, J. M. Provis, Müller cells express the neuronal progenitor cell marker nestin in both differentiated and undifferentiated human foetal retina. *Clin Exp Ophthalmol* **31**, 246–249 (2003).

41. M. M. Emerson, N. Surzenko, J. J. Goetz, J. Trimarchi, C. L. Cepko, Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors. *Dev Cell* **26**, 59–72 (2013).

42. C. Finkbeiner, I. Ortuño-Lizarán, A. Sridhar, M. Hooper, S. Petter, T. A. Reh, 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** (2022).

43. T. Marquardt, R. Ashery-Padan, N. Andrejewski, R. Scardigli, F. Guillemot, P. Gruss, Pax6 Is Required for the Multipotent State of Retinal Progenitor Cells. *Cell* **105**, 43–55 (2001).

44. R. A. Poché, Y. Furuta, M. C. Chaboissier, A. Schedl, R. R. Behringer, Sox9 is expressed in mouse multipotent retinal progenitor cells and functions in Müller Glial cell development. *Journal of Comparative Neurology* **510**, 237–250 (2008).

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

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

47. J. Zhang, C. A. Benavente, J. McEvoy, J. Flores-Otero, L. Ding, X. Chen, A. Ulyanov, G. Wu, M. Wilson, J. Wang, R. Brennan, M. Rusch, A. L. Manning, J. Ma, J. Easton, S. Shurtleff, C. Mullighan, S. Pounds, S. Mukatira, P. Gupta, G. Neale, D. Zhao, C. Lu, R. S. Fulton, L. L. Fulton, X. Hong, D. J. Dooling, K. Ochoa, C. Naeve, N. J. Dyson, E. R. Mardis, A. Bahrami, D. Ellison, R. K. Wilson, J. R. Downing, M. A. Dyer, A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329–334 (2012).

48. P. Blomgren, J. Chandrasekhar, J. A. Di Paolo, W. Fung, G. Geng, C. Ip, R. Jones, J. E. Kropf, E. B. Lansdon, S. Lee, J. R. Lo, S. A. Mitchell, B. Murray, C. Pohlmeyer, A. Schmitt, K. Suekawa-Pirrone, S. Wise, J.-M. Xiong, J. Xu, H. Yu, Z. Zhao, K. S. Currie, Discovery of Lanraplenib (GS-9876): A Once-Daily Spleen Tyrosine Kinase Inhibitor for Autoimmune Diseases. *ACS Med Chem Lett* **11**, 506–513 (2020).

49. K. Kamieniarz-Gdula, N. J. Proudfoot, Transcriptional Control by Premature Termination: A Forgotten Mechanism. *Trends in Genetics* **35**, 553–564 (2019).

50. A. J. Fischer, S. Foster, M. A. Scott, P. Sherwood, Transient expression of LIM-domain transcription factors is coincident with delayed maturation of photoreceptors in the chicken retina. *J Comp Neurol* **506**, 584–603 (2008).

51. S. Irie, R. Sanuki, Y. Muranishi, K. Kato, T. Chaya, T. Furukawa, Rax Homeoprotein Regulates Photoreceptor Cell Maturation and Survival in Association with Crx in the Postnatal Mouse Retina. *Mol Cell Biol* **35**, 2583–2596 (2015).

52. C. M. A. Chen, C. L. Cepko, The chicken RaxL gene plays a role in the initiation of photoreceptor differentiation. *Development* **129**, 5363–5375 (2002).

53. E. Welby, J. Lakowski, V. Di Foggia, D. Budinger, A. Gonzalez-Cordero, A. T. L. Lun, M. Epstein, A. Patel, E. Cuevas, K. Kruczek, A. Naeem, F. Minneci, M. Hubank, D. T. Jones, J. C. Marioni, R. R. Ali, J. C. Sowden, Isolation and

Comparative Transcriptome Analysis of Human Fetal and iPSC-Derived Cone Photoreceptor Cells. *Stem Cell Reports* **9**, 1898–1915 (2017).

54. H. Liu, Y. Zhang, Y.-Y. Zhang, Y.-P. Li, Z.-Q. Hua, C.-J. Zhang, K.-C. Wu, F. Yu, Y. Zhang, J. Su, Z.-B. Jin, Human embryonic stem cell-derived organoid retinoblastoma reveals a cancerous origin. *Proceedings of the National Academy of Sciences* **117**, 33628–33638 (2020).

55. K. Davie, J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Ł. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. Bravo González-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T. Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, N. Konstantinides, M. Fiers, P. Verstreken, S. Aerts, A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982-998.e20 (2018).

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

57. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10 (2011).

58. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).

59. M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).

60. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).

61. S. Orjuela, R. Huang, K. M. Hembach, M. D. Robinson, C. Soneson, ARMOR: An automated reproducible modular workflow for preprocessing and differential analysis of RNA-seq data. *G3: Genes, Genomes, Genetics* **9**, 2089–2096 (2019).

62. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).

63. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

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

65. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–44 (2019).

66. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).

67. K. Alasoo, wiggleplotr: Make read coverage plots from BigWig files. R package version 1.8.0. 1.13.1 [Preprint] (2019). https://github.com/kauralasoo/wiggleplotr.

68. S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008–2017 (2012).

69. P. A. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M. U. Garcia, P. Di Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276–278 (2020).

70. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

71. M. Svoboda, H. R. Frost, G. Bosco, Internal oligo(dT) priming introduces systematic bias in bulk and single-cell RNA sequencing count data. *NAR Genom Bioinform* **4**, lqac035 (2022).

72. M. M. Ibrahim, R. Kramann, genesorteR: Feature Ranking in Clustered Single Cell Data. *bioRxiv*, 676379 (2019).

73. K. Blighe, S. Rana, M. Lewis, EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. [Preprint] (2018). https://github.com/kevinblighe/EnhancedVolcano.

74. Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199–W205 (2019).

75. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).

76. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

77. P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, P. W. Hamilton, QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).

78. U. Schmidt, M. Weigert, C. Broaddus, G. Myers, “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, 2018; http://link.springer.com/10.1007/978-3-030-00934-2_30)vol. 11071 LNCS, pp. 265–273.

79. S. Lee, D. Cobrinik, Improved third-generation lentiviral packaging with pLKO.1C vectors. *Biotechniques* **68**, 349–352 (2020).

80. H. Cheng, Photoreceptor-specific nuclear receptor NR2E3 functions as a transcriptional activator in rod photoreceptors. *Hum Mol Genet* **13**, 1563–1575 (2004).

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

Acknowledgements

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

Funding:

National Institutes of Health grant R01EY026661 (DC)
National Institutes of Health grant R01CA137124 (DC)
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 single-cell 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/. Single-cell RNA-sequencing files and processed final gene and transcript matrices are publicly available at the GEO database under accession number GSE207802.

Supplemental Figures

Figure S1: scRNA-seq summary and cell type identification.

Figure S2: Further characterization of *NRL* isoforms.

Figure S3: Further characterization of *THRB* isoforms.

Figure S4: *RXRG* isoform expression in developing human photoreceptors.

Figure S5: High resolution cluster marker genes and differential RPC versus MG gene expression.

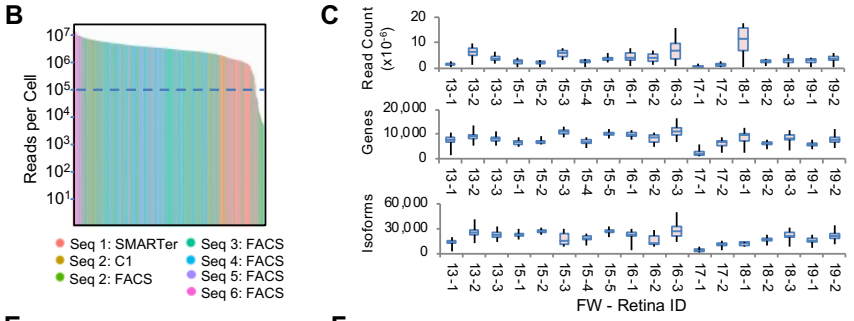
Figure S6: ***NR2E3* and *GNAT2* co-expression in early-born photoreceptors.**

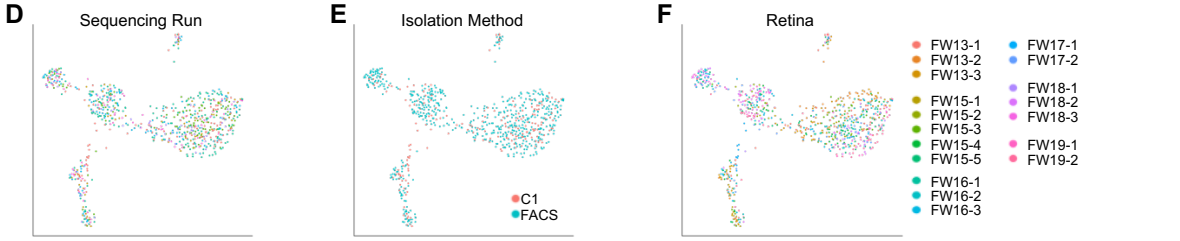
Figure S7: **Additional iPRP bridge region marker genes and regulon activities.**

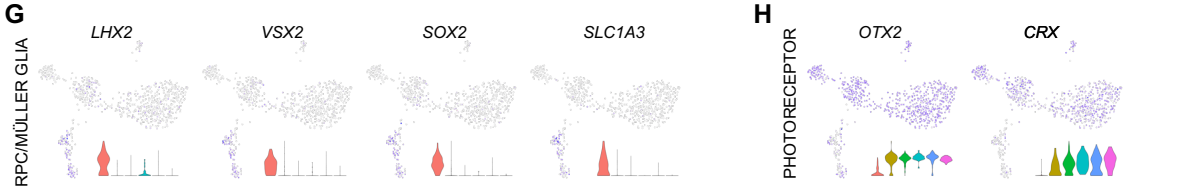
Figure S8: Heatmaps of coregulated gene modules 1-7 across Monocle-derived pseudotime.

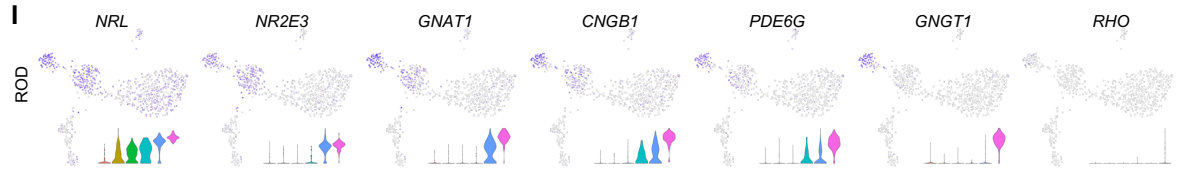
A

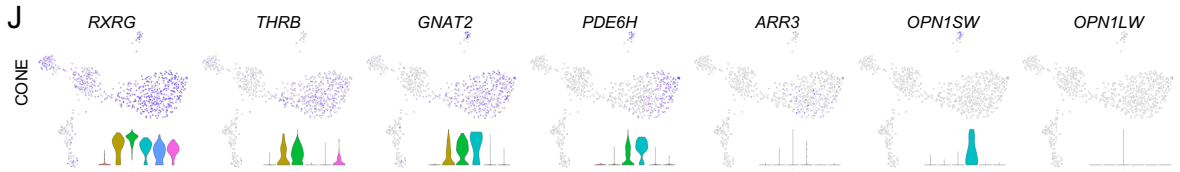
Age (FW)	Number of Retinae	C1 Cells	FACS Cells
13	3	56	127
15	5	53	132
16	3	13	149
17	2	67	-
18	3	-	120
19	2	-	77
Total	18	189	605
	Combined Total	794	

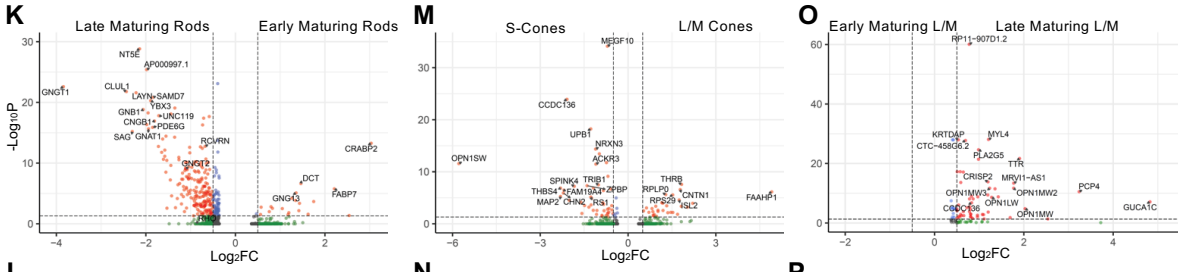












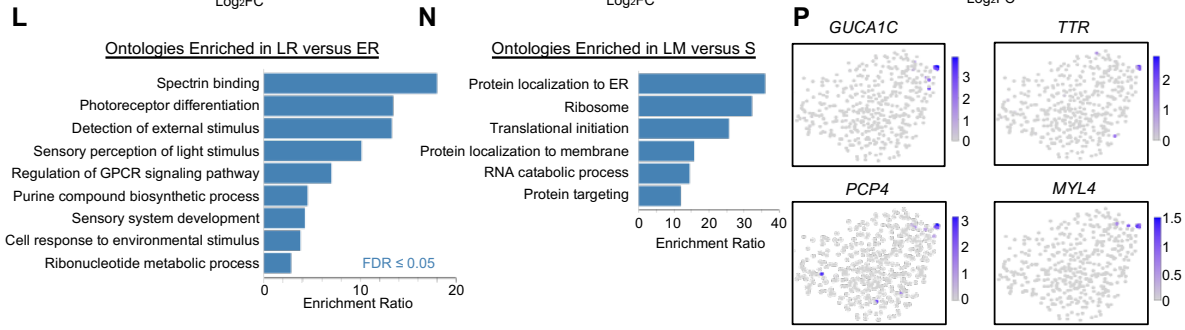
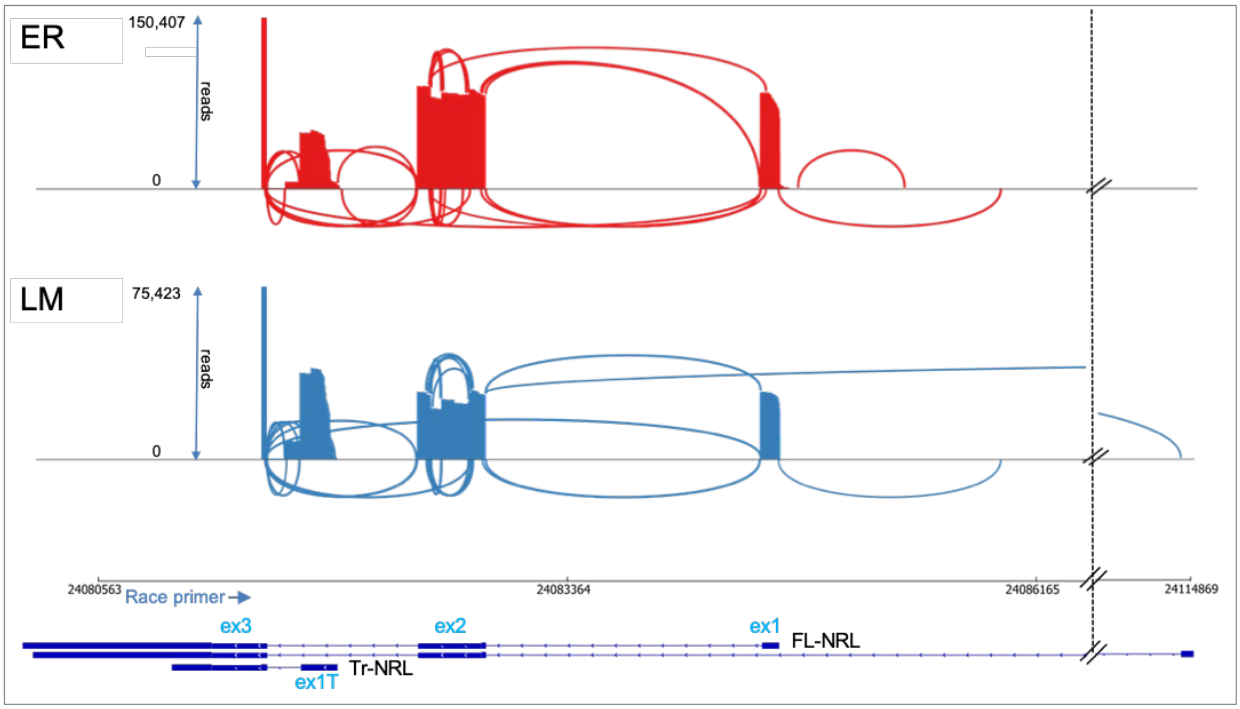
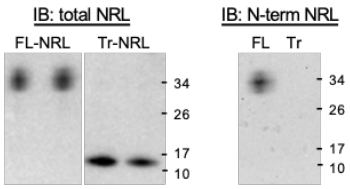


Figure S1: scRNA-seq summary and cell type identification. A. Retinae and cell numbers examined at each age. **B.** Histogram of total read counts for each cell, colored by sequencing run. Seq 1 cells were isolated by C1, Seq 2 cells by FACS or C1 as indicated, and all others by FACS. Dotted line: 100,000 read cutoff for cell exclusion. **C.** Box plots of read counts, genes detected, and transcripts detected per cell ordered by fetal age and specimen number. **D-F.** UMAP plots colored by sequencing run (C), isolation method (D), or retina ID (E). **G-J.** Expression of cell type marker genes for RPCs and Müller glia (G), photoreceptors (H), rods (I), and cones (J). 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/. **K, M, O.** Volcano plots of differential expression (pAdj <0.05, $\log_2$FC >|0.5|) between ER and LR (K), LM and S (M), and spatially separated late maturing L/M cones and remaining LM cluster (O). Labels indicate genes with highest significance and fold change. **L, N.** Overrepresentation of gene ontologies for genes upregulated in LR over ER (GO- Molecular Function instead of Hallmark50) (L) or upregulated in LM over S (N). **P.** UMAP plots of upregulated genes in the late maturing L/M cone group.

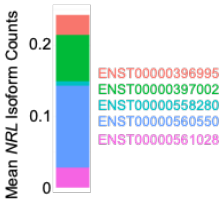
A


ER
150,407
reads
0
LM
75,423
reads
0
24080563
Race primer
24083364
24086165
24114869
ex3
ex2
ex1
FL-NRL
Tr-NRL
ex1T

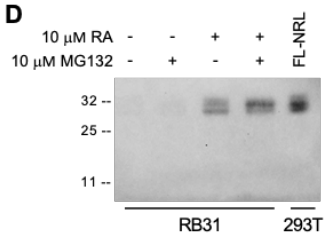
B


IB: total NRL
FL-NRL
Tr-NRL
34
26
17
10
IB: N-term NRL
FL
Tr
34
26
17
10

C


Mean NRL Isoform Counts
0.2
0.1
0
ENST00000396995
ENST00000397002
ENST00000558280
ENST00000560550
ENST00000561028

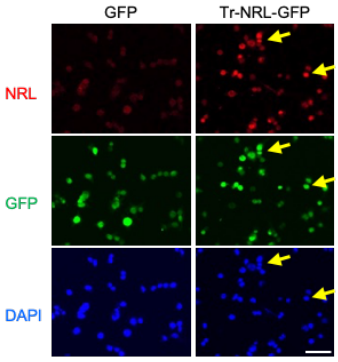
D


10 μM RA - - + +
10 μM MG132 - + - +
FL-NRL
32 --
25 --
11 --
RB31
293T

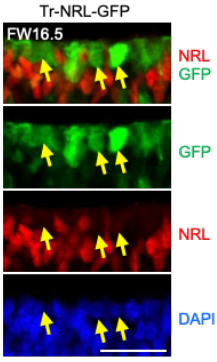
E

pULTRA-EGP-P2A-Tr-NRL (Tr-NRL-GFP)
UBC
EGFP
P2A
Tr-NRL

F


GFP
Tr-NRL-GFP
NRL
GFP
DAPI

G


Tr-NRL-GFP
FW16.5
NRL
GFP
GFP
NRL
DAPI

H

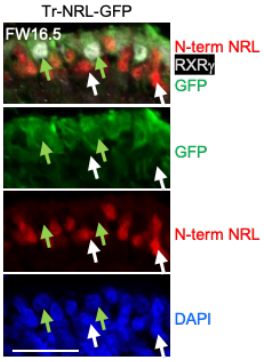

Tr-NRL-GFP
FW16.5
N-term NRL
RXRγ
GFP
GFP
N-term NRL
DAPI

Figure S2: Further characterization of *NRL* isoforms. A. *Top:* Sashimi plots of *NRL* transcript splicing in long read sequencing of 5' RACE reactions from 23 ER and 21 LM cells. *Bottom*: Exon structure of FL and Tr NRL isoforms and RACE primer location. The *NRL* gene is oriented relative to chromosome 14 coordinates. **B.** NRL antibody validation by immunoblot (IB) of ectopic FL-NRL and Tr-NRL expression in HEK293T. *Left:* anti-total NRL. *Right:* anti-N-terminal NRL specific to FL-NRL isoforms. **C.** Average mapped *NRL* isoform expression in scRNA-seq of RB31 retinoblastoma cell line. **D.** 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. **E.** pULTRA-EGFP-P2A-Tr-NRL lentiviral vector used for explanted retina transduction. **F.** NRL Immunofluorescent staining of HEK293T cells 48 h after lentiviral transduction of Tr-NRL-GFP. Scale bar = 50 μm. **G, H.** Immunostaining of total NRL (G) or N-terminal-NRL (H) 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.

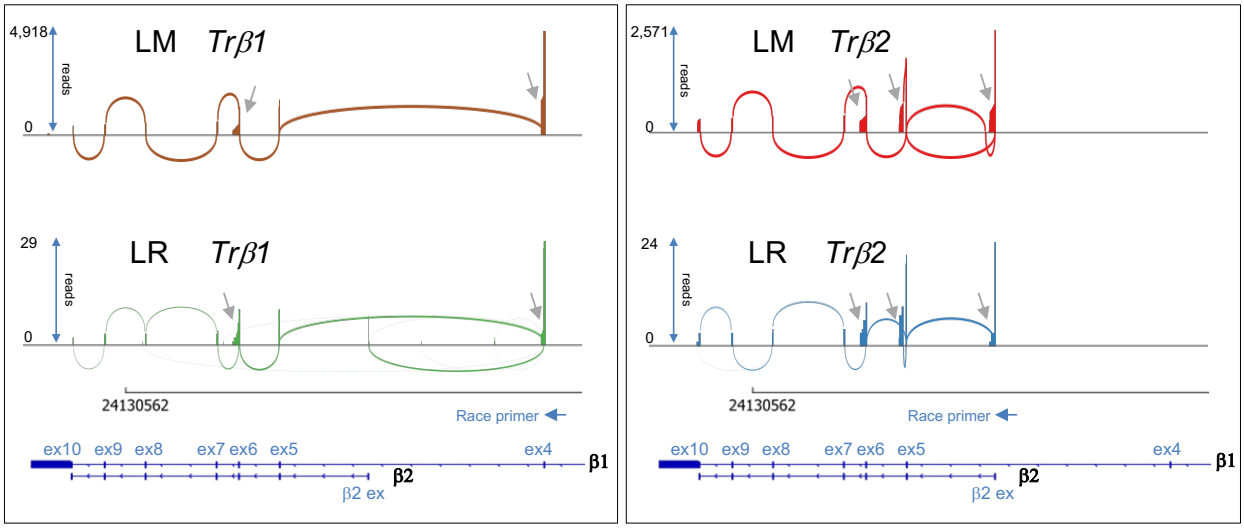
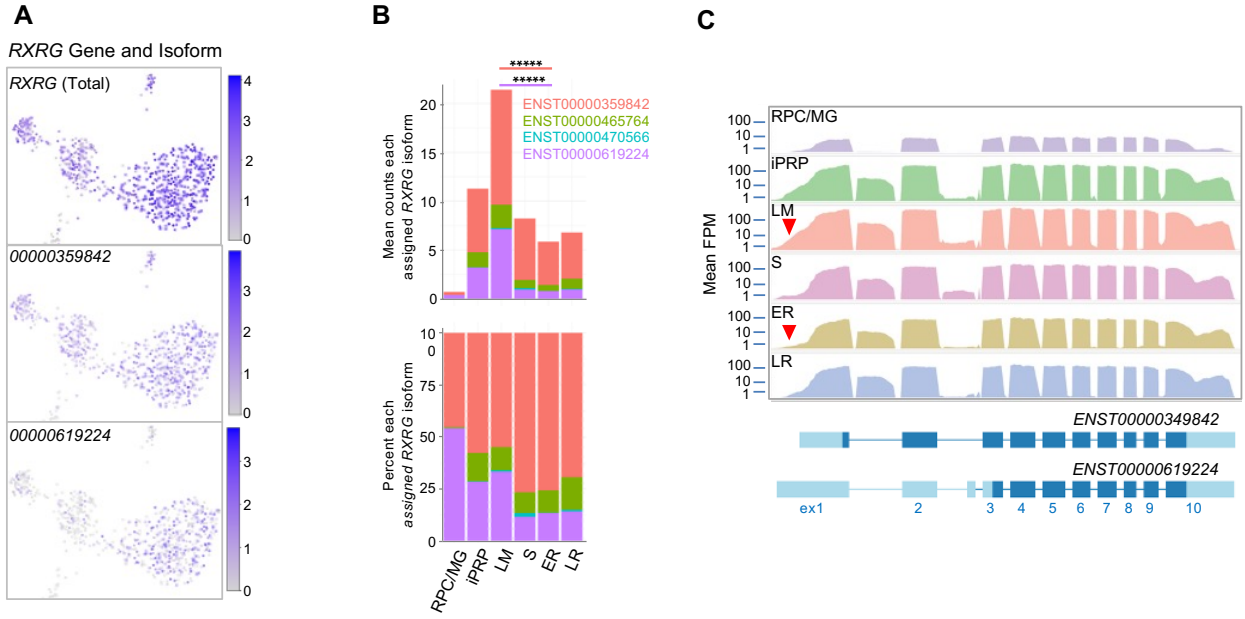
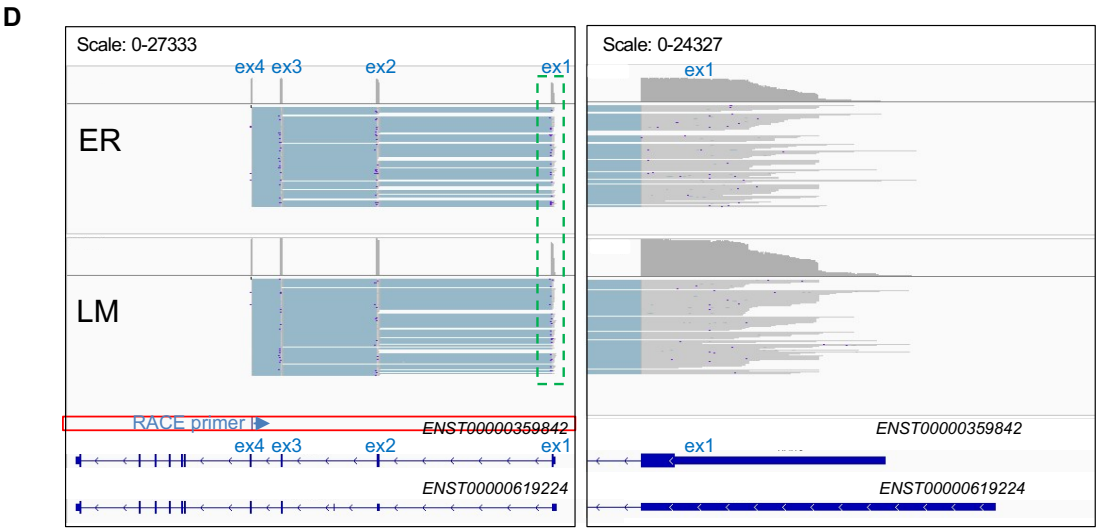


Figure S3: Further characterization of *THRB* isoforms. *Top*: Sashimi plots of *THRB* transcript splicing in long read sequencing of 3' RACE reactions from 23 LM cells and 5 LR cells RACE reactions were initiated from primers specific to TRβ1 exon 4 (*left*) or specific to the unique TRβ2 exon (*right*). 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*: Exon structure of TRβ1 and TRβ2 isoforms and RACE primer locations. The *THRB* gene is oriented relative to chromosome 3 coordinates.


A
RXRG Gene and Isoform
RXRG (Total)
00000359842
00000619224
B
Mean counts each assigned RXRG isoform
Percent each assigned RXRG isoform
ENST00000359842
ENST00000465764
ENST00000470566
ENST00000619224
RPC/MG
iPRP
LM
S
ER
LR
C
Mean FPM
RPC/MG
iPRP
LM
S
ER
LR
ENST00000349842
ENST00000619224
ex1
2
3
4
5
6
7
8
9
10


D
Scale: 0-27333
ex4 ex3
ex2
ex1
ER
LM
RACE primer
ENST00000359842
ENST00000619224
Scale: 0-24327
ex1
ENST00000359842
ENST00000619224

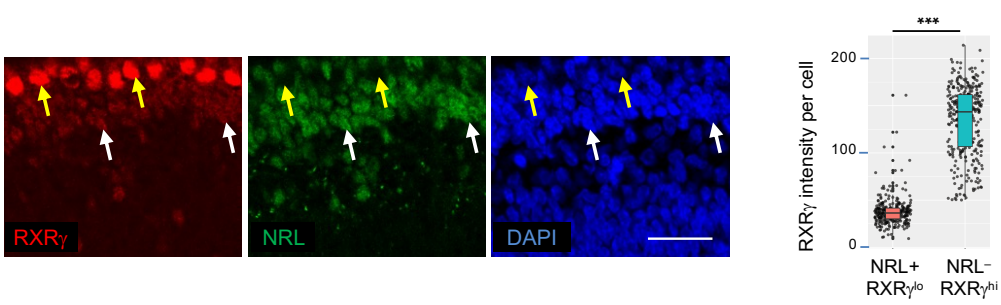
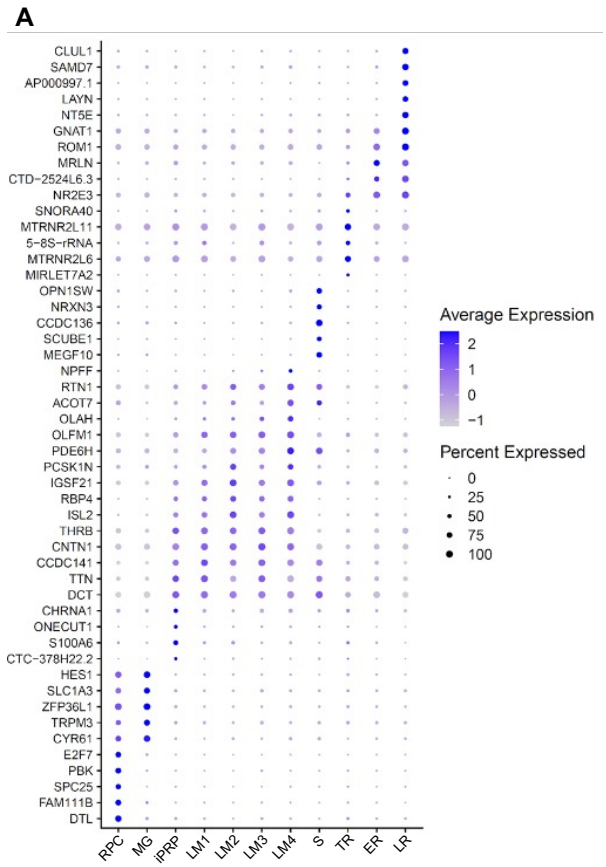
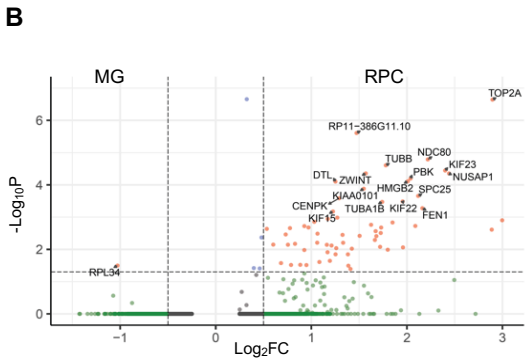
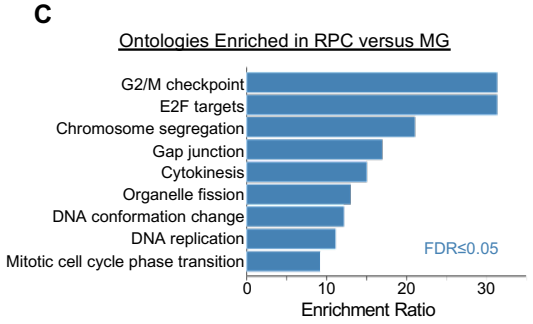

E
RXRγ
NRL
DAPI
RXRγ intensity per cell
NRL+ RXRγlo
NRL− RXRγhi

Figure S4: *RXRG* isoform expression in developing human photoreceptors. A. Expression of *RXRG* and the two most highly assigned *RXRG* isoforms (*ENST00000359842* and *ENST00000619224*). **B.** Average *RXRG* isoform assignments for low resolution clusters presented as total transcript counts (*top*) and as a percentage of the total counts (*bottom*). Lines above figure 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 annotated *RXRG* exons for each cluster. *Bottom*: Transcript structures for *ENST00000359842* and *ENST00000619224.* Red arrowheads: Extended (Ext) *RXRG* 5' UTR in *ENST00000619224.* **D.** Long-read nanopore sequencing of pooled 5' RACE initiated with *RXRG* exon 4 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 transcript structures (*below)*, where expressed sequences are grey and introns light blue. Predicted exons for *ENST00000359842* and *ENST00000619224* isoforms shown below. Blue arrows: *RXRG* exon 4 RACE primer. Panels at *left* show the *RXRG* locus with exon 1 in green box enlarged at right, revealing a similar range of RNA 5' ends in ER and LM populations and no evidence of the unique *ENST00000619224* exon. **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).


A
CLUL1
SAMD7
AP000997.1
LAYN
NT5E
GNAT1
ROM1
MRLN
CTD−2524L6.3
NR2E3
SNORA40
MTRNR2L11
5−8S−rRNA
MTRNR2L6
MIRLET7A2
OPN1SW
NRXN3
CCDC136
SCUBE1
MEGF10
NPFF
RTN1
ACOT7
OLAH
OLFM1
PDE6H
PCSK1N
IGSF21
RBP4
ISL2
THRB
CNTN1
CCDC141
TTN
DCT
CHRNA1
ONECUT1
S100A6
CTC−378H22.2
HES1
SLC1A3
ZFP36L1
TRPM3
CYR61
E2F7
PBK
SPC25
FAM111B
DTL
RPC
MG
iPRP
LM1
LM2
LM3
LM4
S
TR
ER
LR
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
100


B
MG
RPC
TOP2A
RP11−386G11.10
NDC80
TUBB
KIF23
PBK
NUSAP1
DTL
ZWINT
HMGB2
SPC25
KIAA0101
CENPK
TUBA1B
KIF22
KIF15
FEN1
RPL34
-Log10P
Log2FC


C
Ontologies Enriched in RPC versus MG
G2/M checkpoint
E2F targets
Chromosome segregation
Gap junction
Cytokinesis
Organelle fission
DNA conformation change
DNA replication
Mitotic cell cycle phase transition
FDR≤0.05
Enrichment Ratio

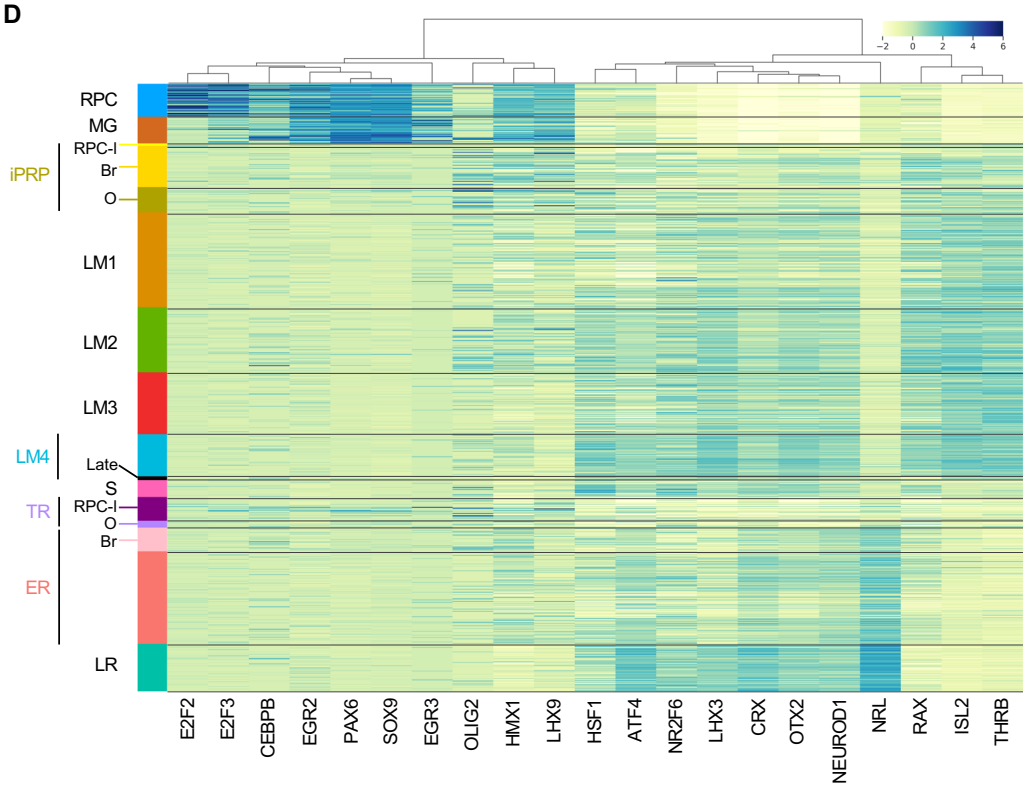

D
RPC
MG
iPRP
RPC-I
Br
O
LM1
LM2
LM3
LM4
Late
S
TR
RPC-I
O
Br
ER
LR
E2F2
E2F3
CEBPB
EGR2
PAX6
SOX9
EGR3
OLIG2
HMX1
LHX9
HSF1
ATF4
NR2F6
LHX3
CRX
OTX2
NEUROD1
NRL
RAX
ISL2
THRB

Figure S5: 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. **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.

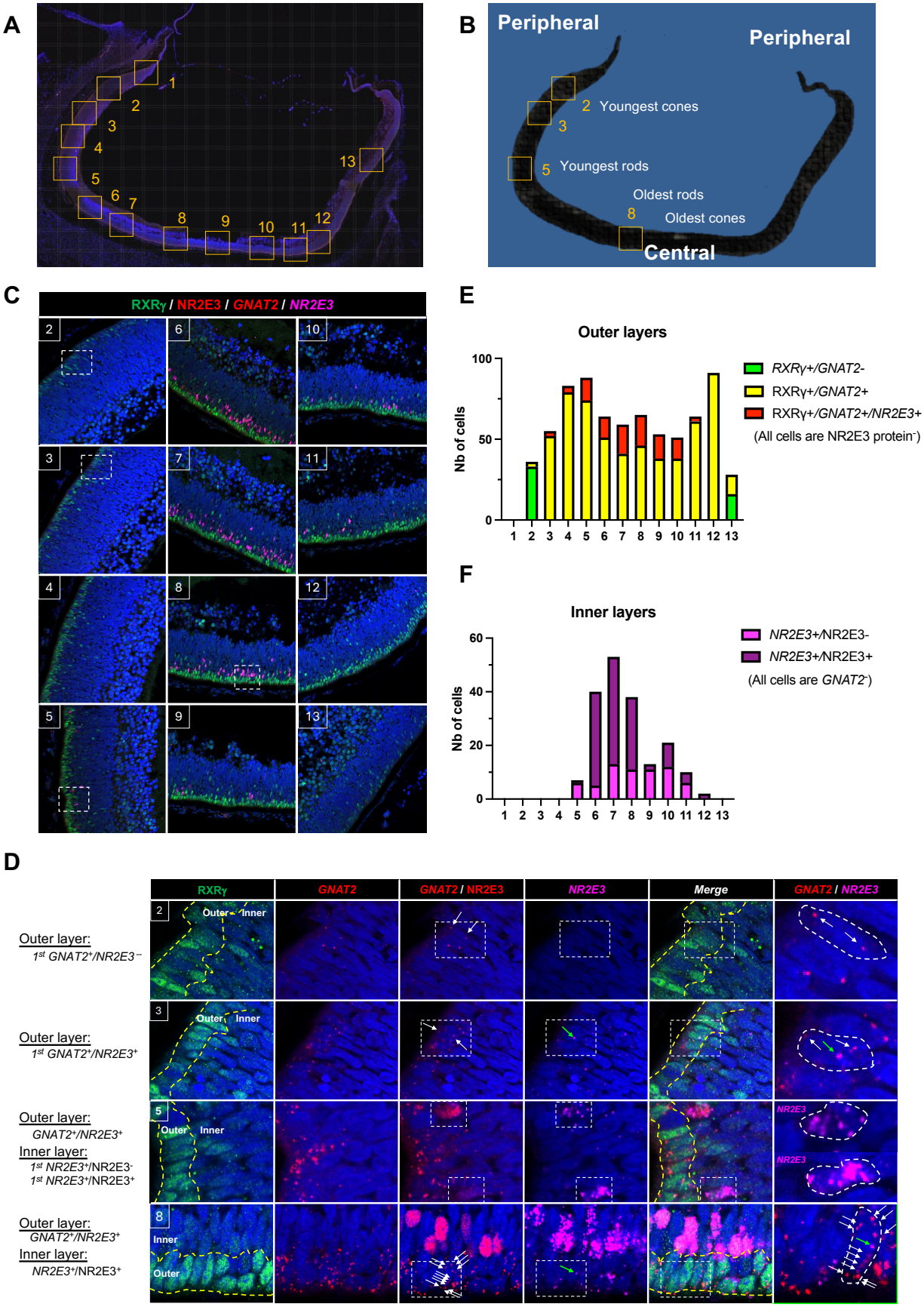

A
1
2
3
4
5
6
7
8
9
10
11
12
13
B
Peripheral
Peripheral
2 Youngest cones
3
5 Youngest rods
Oldest rods
8 Oldest cones
Central
C
RXRγ / NR2E3 / GNAT2 / NR2E3
2
6
10
3
7
11
4
8
12
5
9
13
E
Outer layers
Nb of cells
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13
RXRγ+/GNAT2-
RXRγ+/GNAT2+
RXRγ+/GNAT2+/NR2E3+
(All cells are NR2E3 protein-)
F
Inner layers
Nb of cells
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13
NR2E3+/NR2E3-
NR2E3+/NR2E3+
(All cells are GNAT2-)
D
RXRγ
GNAT2
GNAT2 / NR2E3
NR2E3
Merge
GNAT2 / NR2E3
Outer layer:
1st GNAT2+/NR2E3--
Outer layer:
1st GNAT2+/NR2E3+
Outer layer:
GNAT2+/NR2E3+
Inner layer:
1st NR2E3+/NR2E3-
1st NR2E3+/NR2E3+
Outer layer:
GNAT2+/NR2E3+
Inner layer:
NR2E3+/NR2E3+
Outer
Inner

Figure S6: ***NR2E3* and *GNAT2* co-expression in early-born photoreceptors. A-C.** Combined immunofluorescence staining of RXRγ and NR2E3 and RNA FISH of *GNAT2* and *NR2E3* in FW14 retina. **A.** Tiled composite image of retinal section. Boxed regions are further evaluated in C-F. (B) Diagram of the retina shown in A, indicating regions with the most peripheral (*i.e.*, youngest and least mature) and most central (*i.e.*, oldest and most mature) rods and cones in this section. **C.** Composite images of regions 2-13. Boxes: regions shown at higher magnification in D**. D**. Enlarged images of boxed regions within images 2, 3, 5, and 8. Dotted yellow line demarcates the outer and inner NBL cells analyzed in E, F. White arrows = *GNAT2* puncta. 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 initial expression of *GNAT2* RNA without *NR2E3* RNA in outer NBL nascent cones in region 2, initial co-expression of *GNAT2* and *NR2E3* RNA in early outer NBL cones in region 3, initial expression of *NR2E3* RNA with or without co-expressed NR2E3 protein in region 5, and extensive co-expression of *GNAT2* and *NR2E3* RNA without NR2E3 protein in RXRγ^{hi} outer NBL cones in region 8. **E,F.** Quantitation of (E) outer and (F) inner layer photoreceptor precursors expressing combinations of RXRγ, NR2E3, *GNAT2*, and *NR2E3*.

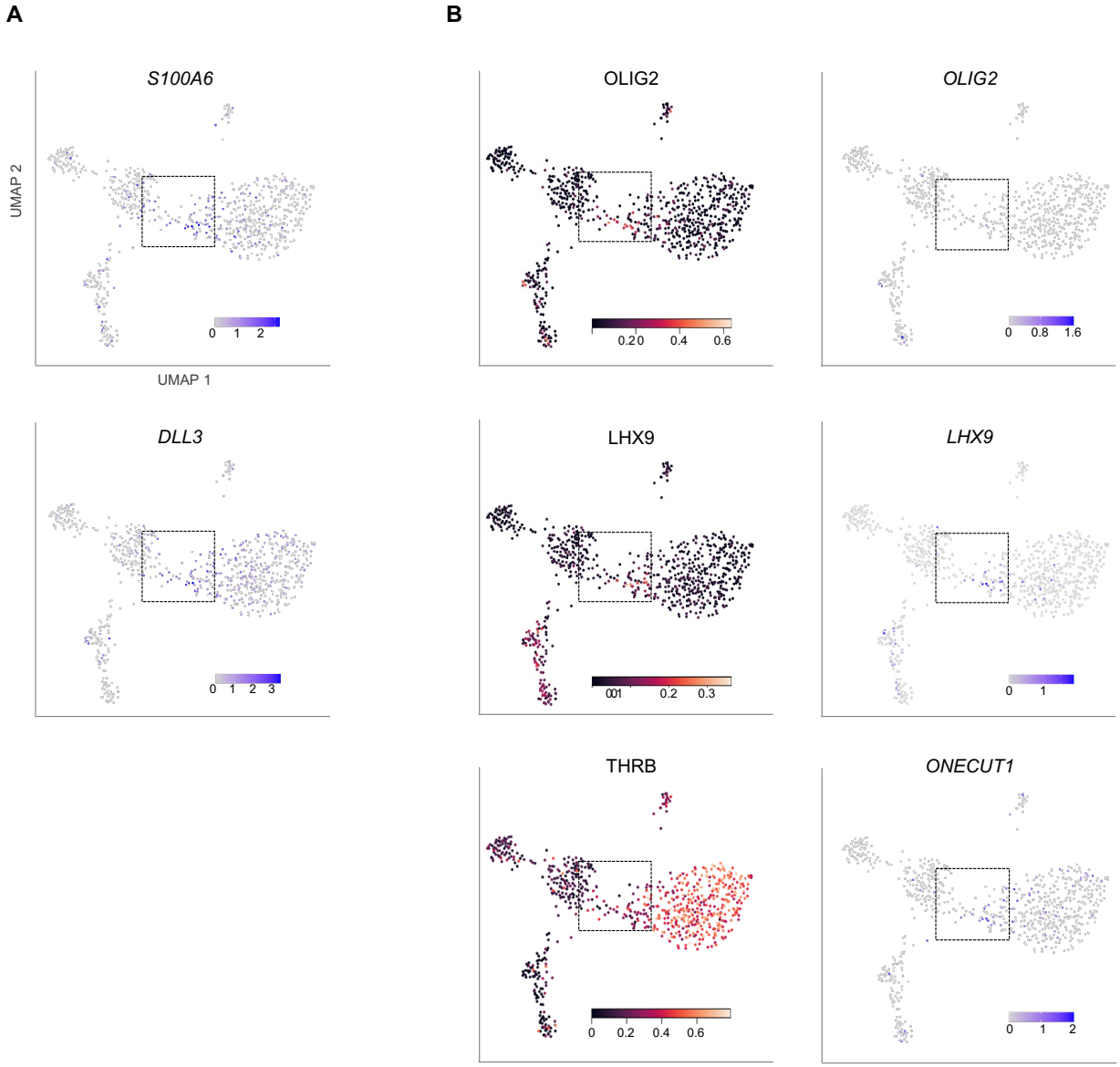


Figure S7: **Additional iPRP bridge region marker genes and regulon activities. A.** UMAP plots of iPRP marker *S100A6* and photoreceptor transition state marker *DLL3* gene expression. **B.** UMAP plots OLIG2, LHX9, and THRB regulon activities (*left*) and *OLIG2*, *LHX9*, and *ONECUT1* gene expression (*right*). Boxes indicate bridge region.

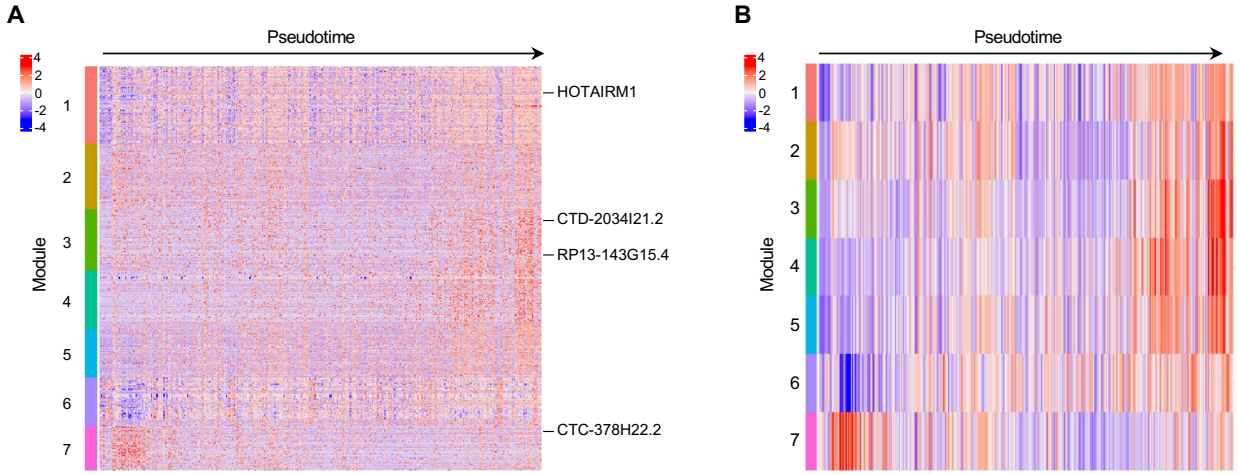


Figure S8: Heatmaps of coregulated gene modules 1-7 across Monocle-derived pseudotime. A. Heatmaps of individual genes in each module across the pseudotime trajectory shown in Figure 6C. lncRNA genes of interest are labeled. **B.** Heatmap of averaged gene module expression across pseudotime.

Supplemental Tables

Table S1: Differentially expressed genes for low resolution rod clusters (LR / ER)

Table S2: Differentially expressed genes for low resolution cone clusters (S / LM)

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

Table S4: SCENIC RSS scores for low resolution clusters

Table S5: SCENIC RSS scores for high resolution clusters

Table S6: Pseudotime-correlated genes and gene modules

Table S7: Differentially expressed genes for low resolution early rod and L/M cone clusters (ER / LM)