Single cell transcriptomics reveals early photoreceptor trajectories and a cancer-predisposed cone precursor state

Dominic W.H. Shayler,[1,2] Kevin Stachelek,[1,3] Sunhye Lee,[1] Jinlun Bai, [1,2] Mark W. Reid, [1] Jennifer G. Aparicio,[1] Yeha Kim,[1] Mitali Singh,[1] Maxwell Bay,[2] Matthew E. Thornton,[4] Brendan H. Grubbs,[4] Michael A. Bonaguidi,[5,6] Cheryl Mae Craft,[7,8] Hardeep P. Singh,[1,7] David Cobrinik*[1,6,7,9]

[1] The Vision Center 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] Maternal-Fetal Medicine Division of the Department of Obstetrics and Gynecology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

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

[6] Department of Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

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

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

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

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

Summary

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 define early photoreceptor developmental states and trajectories that underlie cone-rod differences and identify cone-specific features that contribute to retinoblastoma genesis. The analyses revealed two trajectories from retinal progenitor cells to photoreceptors, one that transitions through a common cone-rod precursor and a second that transitions directly from progenitors to rods. Both L/M cone and rod precursors segregated into early and late maturation states and showed cell type-specific expression of functionally antagonistic *NRL* and *THRB* isoforms. Early cone precursor maturation was characterized by successive expression of four lncRNAs along with *MYCN* and *SYK*, which contributed to the proliferative response to *RB1* loss. These findings reveal previously unrecognized photoreceptor states and trajectories and uncover a role for 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[1]. 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[2], and post-mitotic maturation[3]. However, despite recent advances[4], major events underlying retinal cell fate determination and maturation remain unclear. For example, during photoreceptor development, it remains unclear if fate is fully determined in RPCs or in multi-potent post-mitotic cells. RPC expression of OTX2 and ONECUT1 is thought to control the post-mitotic photoreceptor precursors' expression of fate determinants TRβ2 in presumptive long- or medium-wavelength (L/M) cones and NRL in presumptive rods[5], yet co-expression of TRβ2 and NRL in some precursors[6] suggests that final cell identity is not immediately determined. Following photoreceptor fate commitment, post-mitotic developmental stages have been defined based on morphologic features and phototransduction-related protein or gene expression[7–9], 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 cancer, retinoblastoma.

While much understanding of retinal development derives from study in model organisms[2,10], most models do not recapitulate human retinal features such as a rod-free fovea or tri-color vision[11]. 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[12–14]. While the proliferative response was shown to depend on the human cone precursors' high MDM2 and MYCN expression[13], the developmental basis and

critical features of this cancer-predisposed state have not been defined. Due to the potential importance of species-specific traits, analysis of human tissue is needed to fully define diverse 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[15,16] as well as in human retinae[17–20]. 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[21–23], and gene expression changes associated with the cone fate decision and early development[22,24–26]. However, while high-throughput 3' end-counting methods are invaluable for analyses of large cell numbers in whole tissues, they 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 specify human photoreceptor developmental states have not been interrogated 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 RNA velocity, pseudotemporal trajectory reconstruction, and single-cell regulatory network inference and clustering (SCENIC) to interrogate individual cell states. These data enabled us to discriminate photoreceptor developmental states, identify transcript isoform-specific expression patterns, elucidate post-mitotic developmental trajectories, and define retinoblastoma cell of origin features that contribute to retinoblastoma genesis.

Results

Full-length scRNA-seq identifies 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 based on forward-scatter, side-scatter, and CD133/CD44/CD49b staining 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[13] 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, or horizontal cells (*POU4F1*, *POU4F2*, *POU4F3*, *TFAP1A*, *TFAP1B*, *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 1B, S1A,B). Sequencing batches were normalized and transcriptomes clustered and visualized in uniform manifold approximation and projection (UMAP) plots that integrated cells across different retinae, single cell isolation methods, and sequencing runs (Figure 1C, S1C-E).

Low resolution Louvain clustering generated six clusters that segregated into mostly distinct UMAP domains (Figure 1C). 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 1D, S1F-I). Among these, 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 1C,D).

Two clusters expressing the rod determinant *NR2E3* were designated early maturing rod (ER) and late maturing rod (LR) (Figure 1C), based on the latter's increased expression of rod phototransduction genes *GNAT1*, *CNGB1*, *PDE6G*, *GNGT1* and *RHO* (Figure 1D, S1H). Differential expression analysis revealed other genes upregulated in late rods (*GNB1, SAMD7*, *NT5E*)[27,28] as well as the downregulated *CRABP2*, *DCT*, and *FABP7* (Figure S1J, Table S1). Genes upregulated in the LR cluster were enriched for photoreceptor and light sensing gene ontologies (Figure S1K) including spectrin binding, likely relating to proteins that control photoreceptor polarity and synapse formation[29,30], and purine biosynthesis and ribonucleotide metabolic processes, potentially related to the developing photoreceptors' high NAD+ requirements[31].

Cones segregated into distinct S- and L/M-cone clusters, with differential expression of cone subtype markers (*OPN1SW*, *THRB*), recently identified S-cone enriched genes (*CCDC136*, *UPB1*)[32–34]**,** novel S cone-upregulated genes (*MEGF10, NRXN3*, *ACKR3*), and the L/M cone transcription factor *ISL2*[22], among others (Figure 1C,D, S1L, 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 S1M) due to increased expression of ribosomal protein genes *RPL23A*, *RPLP0*, *RPS19*, *RPS27*, *RPS27A*, *RPS29*, *RPS3A*.

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 and late maturing L/M cones[35,36], respectively, based on the latter's increased expression of cone phototransduction genes *OPN1LW* (encoding L-opsin), *PDE6H*, and *GUCA1C* (Figure 1D, S1N,O), analogous to *RHO*, *PDE6G*, and *GNGT1* upregulation in late rods. Differential gene expression analysis against the early L/M population also showed upregulation of the three M-opsin genes (*OPN1MW*, *OPN1MW2*, *OPN1MW3*)*, TTR*, encoding the retinol-binding protein transthyretin, *PCP4*, a small peptide that binds calmodulin previously noted in foveal cones[37], and *MYL4*, a myosin light chain gene previously noted as upregulated in retinal organoid L/M cones[34], among

others (Figure S1N,O, Table S3). The low proportions of *OPN1MW*+ and *OPN1LW*+ late maturing L/M cones is consistent with a prior analysis of similar-age retinae[22] and the further upregulation of these proteins in later maturation[38].

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[39]. 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 1E-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 1E). RAX activity also decreased in the 5-cell late maturing L/M cone group ($p < 0.01$), 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, likely related to the transitional nature of this population. However, SCENIC did not identify S cone-specific regulons, in keeping with the notion that S cones represent a default photoreceptor state influenced by pan-photoreceptor factors such as OTX2, CRX, and NEUROD1[5]. Thus, deep, full-length scRNA-seq enabled identification of regulons underlying RPC and developing photoreceptor states at the single cell level.

Alternative *NRL*, *RXRG*, and *THRB* isoforms in rod and cone populations

Although cone and rod precursors segregated into distinct clusters, mRNAs encoding rod-determining factor NRL, cone marker RXRγ, and L/M cone determining factor TRβ were unexpectedly co-expressed in both rod and cone populations, with mean *NRL* expression 3.8-fold higher in the ER vs LM cluster, mean *RXRG* expression 4.7-fold higher in LM vs ER, and mean *THRB* expression 2.5-fold higher in LM vs ER (Figure 2A, S1H,I). Cone *NRL* expression

was unexpected given NRL's ability to induce rod differentiation[34] and the rod-specific NRL regulon activity (Figure 1F). Similarly, rod *RXRG* and *THRB* expression were unexpected given their roles in cone gene expression and differentiation[40–42] and the L/M cone-specific THRB activity (Figure 1F). However, interrogation of full-length scRNA-seq Ensembl isoform assignments and exon coverage revealed cell type-specific regulation of transcript isoforms concordant with higher NRL protein in rods and TRβ2 and RXRγ protein in cones.

For *NRL*, the most highly assigned isoform (*ENST00000397002*) encoded full-length NRL (FL-NRL) (RefSeq NP_001341697.1), while the second-most assigned isoform (*ENST00000560550*) used an alternative "P2" promoter and first exon and encoded an N-terminally truncated NRL (Tr-NRL) retaining the leucine zipper DNA binding domain but lacking the minimal transactivation domain[43] (Figure 2B,C). While both isoforms increased in the ER vs LM cluster, the FL-NRL-encoding *ENST00000397002* showed a higher 7.0-fold increase and the average *ENST00000397002:ENST00000560550* ratio was 2.8:1 in early rods and 0.72:1 in L/M cones, suggesting cell type-specific regulation (Figure 2B). As Ensembl isoform assignments are not definitive, we further assessed relative expression of the unique first exon sequences. Consistent with the assigned isoform ratios, mean Tr-NRL first exon coverage was higher in S and L/M cones while FL-NRL first exon coverage was higher in rods (Figure 2C). To assess first exon usage in individual cells, we compared the difference in the number of reads mapped to each first exon relative to the total reads of both. This further confirmed that the Tr-NRL first exon predominated in developing cones whereas the FL-NRL first exon predominated in rods (Figure 2D). Notably, iPRP cells had an evenly distributed range of FL-NRL and Tr-NRL first exon preferences consistent with the bipotential nature of this population.

While the Tr-NRL isoform was not previously described, a truncated *NRL* isoform with alternative FL-NRL exon 1 splicing, termed DD10 (approximated by RefSeq NM_001354770.2), was identified in adult retina[44] (Figure 2C). However, we detected fewer reads spanning the unique DD10 splice junction compared to the unique Tr-NRL junction (5,942 versus 57,048) in

our dataset. In keeping with the DD10 ability to interfere with FL-NRL transactivation[45], ectopic Tr-NRL suppressed FL-NRL transactivation of a *PDE6B* promoter in NIH-3T3 cells (Figure 2E, S2A,B). Thus, the novel cone-enriched Tr-NRL antagonized FL-NRL-mediated transactivation.

Despite our detection of *NRL* RNA, NRL protein has not been detected in cones. To assess endogenous Tr-NRL expression, we performed immunoblot analysis on CHLA-VC-RB-31 retinoblastoma cells[46], which expressed FL-NRL- and Tr-NRL-encoding transcripts in a cone-like 0.57:1 ratio (Figure S2C). As in other retinoblastoma cells[47], 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 retina with an EGFP-P2A-Tr-NRL cassette (Figure S2E-H). These findings suggest that Tr-NRL is poorly translated or unstable in developing cone precursors.

For *RXRG*, *ENST00000359842* was imputed as the most highly expressed isoform with a 2.5-fold higher mean expression in L/M cones than in early rods. The second-most assigned isoform, *ENST00000619224*, uses an alternative promoter and was inferred to have 7.5-fold higher expression in L/M cones (Figures 2F, S2I-J). Concordantly, the alternative promoter's unique 5' UTR sequences had 10.8-fold increased mean counts in LM versus ER cells ($p < 2.2e-16$, Welch's T-test). These observations show that an alternative *RXRG* promoter is upregulated in cones and likely contributes to their higher RXRγ expression (Figure S2K).

THRB expression in L/M cones was largely attributed to isoforms encoding the L/M cone-specific TRβ2 (*ENST00000280696)* and the widely expressed TRβ1 (*ENST00000396671)* (Figure S2L,M). Notably, LR cells had ~50% mean *THRB* expression compared to LM cells, but preferentially expressed TRβ1 exons 1-4, and to a lesser extent exons 5-6 (Figure 2G). Most reads mapping to these exons were assigned to uncharacterized Ensembl isoforms (*ENST000004228492* and *ENST00000431815*) but likely reflect novel isoforms that retain *THRB* exon 4 and exon 6 splice donor junctions and terminate in the subsequent introns

(Figures 2G,H, S2M,N). Interestingly, transcripts extending through the exon 6 splice donor are predicted to encode a truncated TRβ1 protein retaining the A/B and DNA binding domains but lacking the ligand binding and ligand-dependent transactivation domains (Figure 2G). As this truncated structure is similar to TRβ4, which antagonizes thyroid hormone-induced transcription[48], we infer that the Exon1-6 isoform may compete with full-length TRβ1. Thus, our full length transcriptome dataset enabled identification of potentially antagonistic truncated *THRB* isoforms in rods as well as antagonistic truncated *NRL* isoforms in cones.

Two post-mitotic photoreceptor precursor populations and trajectories

To further define photoreceptor developmental states and trajectories, 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[49] (Figure 3A,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 marker gene expression, and yielded a new cluster, here designated transition rods (TR), derived from the low resolution iPRP and rod clusters (Figure 3A, 3C, S3A).

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 3D, S3A,B). The MG cluster was spatially segregated, lacked 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*[15,50,51] (Figure 3C,D, S3A). Concordantly, differential gene expression analyses revealed upregulation of cell cycle-related genes and G2/M checkpoint and E2F target ontologies (Figure S3B,C) in the RPC cluster but no significantly upregulated genes or ontologies in the MG population, consistent with early MGs resembling quiescent RPCs[52].

High resolution clustering also resolved iPRP and TR cells positioned near the RPCs in UMAP space (Figure 3C). RNA velocity suggested that these 'RPC localized' cell groups flowed towards the larger iPRP population and towards early maturing rods, respectively. Compared to RPCs and MGs, RPC-localized iPRP and TR cells had minimal expression of cell cycle-related cyclin RNAs or RPC markers *PAX6*, *LHX2*, and *VSX2*, and had increased expression of photoreceptor determinants *OTX2* and *CRX* (Figure 3D), consistent with their being immediately post-mitotic photoreceptor precursors. The iPRP cells in this region also showed upregulation of the early cell fate determinant *ONECUT1*[53,54], the L/M cone determinant *THRB*, and the cone differentiation marker *GNAT2,* but lacked the rod determinant *NR2E3*, indicative of a cone-bias (Figure 3D). In contrast, TR cells in the same region trended towards higher expression of the rod-related *NRL* and *NR2E3* ($p= 0.19$, $p=0.054$, respectively).

SCENIC regulons further clarified the identities of the four RPC-localized clusters (Figure 3E). The MG cluster was best specified by PAX6 and SOX9, both described in RPCs and MGs[50,55,56], and by early response gene EGR3, 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 increased *THRB* and *NRL* regulon activities, with iPRP cells having greater *THRB* regulon signal than TR cells, and TR cells having higher NRL regulon signal ($p<0.05$ for both), consistent with the bifurcated direction of gene expression change implied by RNA velocity (Fig. 3C,E). 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[22,23,26].

A common cone-rod photoreceptor precursor state

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 4A). RNA velocity vectors in this bridge region bifurcated towards rods or cones, with individual cells expressing cone markers (*GNAT2*, *THRB*), rod markers (*GNAT1*, *NR2E3*), or, in many cases, both (Figure 4A,B). 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 neighboring ER regions, cells with NRL and THRB regulon signals intermixed, and some cells showed strong signals for both (Figure 4C, arrows). While cells with both signals were also present in other clusters, bridge region iPRP cells had more similar and lower NRL and THRB regulon signals, with small mean absolute differences between NRL and THRB regulon z-scores (Figure 4D). The mean difference between regulon signals increased in rod and cone clusters outside the bridge as NRL or THRB regulons became more dominant (Figure 4D). Interestingly, TR and S clusters also had small differences between NRL and THRB regulon z-scores and low THRB and NRL activities consistent with a lack of NRL and THRB activities in S cones and a not-yet realized NRL activity in transitional rods. The iPRP bridge region cells' co-expression of THRB and NRL regulons and bifurcating RNA velocity vectors are consistent with the existence of a common cone-rod precursor[6] in which fate commitment is accompanied by increased THRB and or NRL activity.

To further assess whether iPRP bridge cells comprise a distinct subpopulation, we sought to identify RNA markers of this region. Among four iPRP marker genes (Figure S3A), *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 tight zone of *THRB+, GNAT2+* cone precursors, and *S100A6* was more widely expressed (Figure 4E, S4A). Bridge region cells also expressed *ATOH7* and *DLL3*, which were previously

proposed to define transitional photoreceptor precursors and promote cone fate determination[21,23,57,58]. *ATOH7* was largely restricted to cone-directed cells whereas *DLL3* was more broadly expressed (Figure 4E, S4A).

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* expression 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 4F-H). The expression of *CHRNA1* in the earliest born rods and cones is consistent with their derivation from a common precursor.

To identify factors that regulate transcriptomic states during iPRP bridge cell fate determination, we examined the distribution of the most iPRP-specific regulons, OLIG2 and LHX9 (Figure 1E, Table S4). As Olig2 was previously detected in mouse RPCs in which Onecut1 enabled *Thrb* expression and cone fate determination[53,59], the OLIG2 regulon was expected to be most active in RPCs and early, immediately post-mitotic iPRPs. Instead, both the OLIG2 and LHX9 regulons showed focal signal in *ONECUT1*+ cone-directed iPRP cells distal to the bridge region and immediately prior to THRB regulon upregulation (Figure S4B). 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 changing morphology, outer segment formation, and phototransduction pathway upregulation[8,9]. 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 (Figure S5A). Marker gene expression differences were subtle and best distinguished by

sequential increasing expression of *ACOT7*, *RTN1*, *PDE6H*, *OLAH*, and *NPFF* (Figure S3A, S5B). However, among the three regulons with highest LM1-4 specificity scores, THRB and ISL2 showed only small differences between clusters while LHX3 increased across LM1-4 cones as well as across early and late rods (Figure S5C). Overall, the lack of cluster-specific marker genes and regulons suggest that LM1-LM4 denote different stages of a graded maturation process rather than distinct transcriptomic states.

We further evaluated gene expression changes across maturing L/M cones through trajectory analyses. RNA velocity suggested that two trajectories flowed from immature photoreceptor precursors towards the most mature cells in LM4 (Figure S5A). One arm (along the top of the UMAP space) appeared to progress directly from the iPRP cluster towards LM4, whereas the other entered a central low velocity region comprised of LM1, LM2, and LM3. These trajectories were distinguished by fetal age, with the upper route primarily comprised of FW13 cells and the lower route primarily FW15-19 (Figure S5D), although the paths were not sufficiently distinct to enable differential gene expression analysis. We then used Monocle 3[60] to pseudotemporally order the iPRP and L/M cones and identify pseudotime-correlated genes and coregulated gene modules. This produced a single trajectory rather than several age-dependent paths suggested by RNA velocity (Figure 5A), likely due to the transcriptomic similarity in this population. Monocle identified 799 pseudotime-correlated genes (q-value < 0.05, expression > 0.5 in >5% of ordered cells) (Table S5), which composed nine gene modules characterized by early expression (cluster 4), initially low and rising expression (clusters 2,5), intermittent expression that increased late (cluster 1), and late expression (clusters 3, 6-9) (Figure 5B). Only module 2 genes significantly over-represented gene ontologies, specifically, drug metabolism, ribosome, light sensing, and MYC targets.

Among the top 15 pseudotemporally-correlated genes in each temporal expression group, we identified four lncRNAs that were sequentially expressed across iPRPs and maturing L/M cones (Figure 5C), 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 5D,E). 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* (blue), peripheral cones with high *HOTAIRM1* only (green), cones where all four lncRNAs show low expression (red), and parafoveal/foveal cones with upregulated *RP13143G15.4*, *CTD-2034I21.2*, and *CTC-378H22.2* (purple) (Figure 5F). Except for central retina *CTC-378H22.2*, these data validate the patterned expression of lncRNAs implied by the UMAP plot and suggest that specific lncRNA combinations mark maturing L/M cone stages.

Cone intrinsic MYCN and SYK contribute 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[13,14]. 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 6A, Table S6). Among cone-enriched genes, the top three enriched ontologies related to translation initiation, protein localization to membrane, and MYC targets (Figure 6B). Although *MYC* RNA was not detected, the upregulation of MYC target genes was of interest given the cone precursors' high MYCN protein expression (Figure 6C) and the critical role of MYCN in the cone precursor proliferative response to pRB loss[12,13]. Concordantly, the LM cluster had increased *MYCN* RNA ($\log_2FC = 0.54$) and MYCN regulon activity, representing the eighth highest LM cluster regulon specificity score (Figure 6D,E).

Among other differentially expressed genes, we noted cone-specific upregulation of *SYK* (Figure 6A,F, Table S6), which encodes a non-receptor tyrosine kinase that acts upstream of MYC[61]. Whereas SYK was previously implicated in retinoblastoma genesis and proposed to be

induced in response to pRB loss[62], its expression was not previously reported in developing fetal retina. Indeed, 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 6G). The loss of SYK expression with advanced cone maturation is consistent with the lack of SYK in cones of normal retina adjacent to a retinoblastoma tumor[62].

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[13], treated with the specific SYK inhibitor GS-9876[63] for 12 days, and examined for cone precursor cell cycle entry by co-immunocytochemical staining for RXRγ and Ki67. Strikingly, GS-9876 treatment suppressed the proliferative response to pRB knockdown at all concentrations from 1.0 – 5.0 μM (Figure 6H). These data reveal that cone precursor intrinsic SYK expression and activity are critical to the proliferative response to pRB loss (Figure 6I).

Discussion

This study evaluated transcriptome state changes associated with human cone and rod photoreceptor development using deep full-length RNA-sequencing. Whereas prior scRNA-seq studies provided insights into RPC fate determination and transcriptomic trajectories using 3' end counting[21–25,32,64–66], full-length sequencing enabled 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 developing cones, a rare population with unique cancer cell of origin properties[12–14]. The resulting dataset provides a rich repository for photoreceptor development and disease research.

One advantage of full-length scRNA-seq is that it enables detection of transcript isoforms in individual cells. Among established photoreceptor markers, NRL, TRβ2, and RXRγ

are described as cell type-specific and fate-determining[2,42,67], yet at the RNA level each was expressed in both rod and cone precursors (Figure 2, S2). Our data revealed that cones preferentially express a novel *NRL* transcript predicted to encode a truncated NRL (Tr-NRL) lacking a transactivation domain. As with a previously described *NRL* isoform[45], detected here at lower levels, Tr-NRL opposed transactivation by full-length NRL and thus may suppress rod-related transcription. However, our inability to detect intrinsic Tr-NRL or to overexpress Tr-NRL in retinal cone precursors suggests there are additional layers of regulation, possible effects at the RNA level, and specific contexts in which these isoforms act. Similarly, late maturing rods expressed truncated *THRB* transcripts lacking the ligand binding and ligand-dependent transactivation domains; as one such isoform resembles the antagonistic TRβ4[48], it likely opposes cone-related gene expression. The identification of these alternative isoforms opens new opportunities to understand rod and cone fate commitment. More generally, the publicly available data and analysis tools may be mined to characterize alternative isoforms of any gene expressed during human photoreceptor development.

Understanding human photoreceptor development requires the identification of all 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[39]. Our deep sequencing approach uncovered discrete states distinguished by gene expression and regulon activity. For example, low resolution Louvain clustering resolved early and late maturing rod populations distinguished by the latter's unique expression of *GNGT1* and *RHO*, increased expression of other genes (*e.g.*, *GNAT1*, *PDE6G*) and the NRL, ATF4, CRX, and LHX3 regulons, and decreased expression of the HMX1 and RAX regulons. In contrast, L/M cones formed a single cluster characterized by high THRB and ISL2 regulon activity – consistent with THRB driving L/M cone identity[40,42] and with cone precursor *ISL2* expression[22,68] – and a distinct late maturing cone group with upregulated L/M opsin gene expression and decreased RAX

regulon activity. Downregulation of the RAX regulon marked a late maturation state change in rod and cone precursors, consistent with decreasing RAX protein during photoreceptor maturation and with RAX-mediated suppression of the cone opsin and *rhodopsin* promoters[69,70].

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 (Figure S5C), consistent with LM1-4 comprising different stages of a graded maturation process. Early cone states have also been parsed in mice[24,25], although the identified mouse cone cluster or 'wave' marker genes were either diffusely or minimally expressed in our human cone development trajectory. RNA velocity suggested that L/M cones follow subtly different early maturation trajectories according to age, potentially related to differences between early born foveal and later born peripheral cones[8,33,66]. More intriguingly, pseudotemporal analysis identified a maturation trajectory with successive expression of L/M cone-specific lncRNAs that was validated in developing retinal tissue. Notably, *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 demonstrating sequential expression[22,25,71]. The sequential expression of these lncRNAs in the absence of discrete regulon changes implies they demarcate early L/M cone precursor substates, yet it remains unclear if they play a developmental role or are coincidental signposts.

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 showed higher cone-like or rod-like gene expression and regulon activities, respectively, and had RNA velocities toward distinct cone/rod hybrid or rod populations. Prior scRNA-seq analyses have not to our knowledge shown evidence of divergent routes to photoreceptors, although UMAP plots showed spatially distributed cone and rod neurogenic precursors that could comprise similar populations[22,23]. The iPRP versus TR decision appears to be made in

RPCs or rapidly in the post-mitotic period, based on the biased THRB versus NRL regulon activity in the RPC-adjacent 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 before diverging towards definitive rod and cone states. This population resembled the T3 transition state identified via 3' end-counting[23], with expression of *ATOH7*, *DLL3*, *OTX2*, and the new marker *CHRNA1*. Bridge population 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[53] their expression in the iPRP population suggests a similar role in bipotential photoreceptor precursors. Previous work suggested that final photoreceptor fate can be determined either pre-mitotically[53] or post-mitotically[6,72]; the current data suggests that both may be correct in that early iPRP cells are yet to decide between photoreceptor fates while early TR cells are committed to produce rods. Notably, the rod-directed bridge and TR populations imply there are two pathways to rods; while in keeping with other examples of diverging and then re-converging trajectories[73], this notion requires lineage tracing studies for formal support. If verified, a possible explanation for these routes may be that RPCs shift from iPRP generating in early retinal development to TR generating in later development, when large numbers of rods must be produced.

Gene expression differences in early cone versus rod photoreceptors were also mined to identify factors that impact human cone cell sensitivity to *RB1* loss[13]. We detected significant upregulation of genes targeted by MYC along with upregulated MYCN protein[12], which shares target genes with MYC during cell cycle entry from quiescence[74], along with upregulated *MYCN* RNA and regulon activity. As MYCN is required for the proliferative response to pRB loss[13] and triggers retinoblastoma when amplified and overexpressed[75,76], these findings suggest that increased intrinsic *MYCN* RNA expression contributes to cone precursor proliferation and retinoblastoma tumorigenesis. We further detected the cone precursors' significantly increased

SYK RNA and dramatically increased protein. SYK was previously detected in human retinoblastoma and in *RB1*-null retinal organoid tumor-like masses but not in healthy tumor-associated retina[62,77]. Our finding that SYK is highly expressed in FW16 but not FW18 foveal cone photoreceptors implies that its expression in retinoblastoma reflects a cell-of-origin feature that is lost in later cone development. SYK inhibition impaired cone cell cycle entry after pRB knockdown, indicating that native SYK expression, rather than *de novo* induction, contributes to the cone precursors' initial proliferation and retinoblastoma tumorigenesis (Figure 6H)[14].

In conclusion, 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 and associated RNA velocities suggested novel cell fate determination trajectories and revealed lncRNA patterning of early maturing L/M cone sub-states. It is unclear if high-throughput 3' end counting would enable similar cell state discrimination, or if detection of low-expressed genes via deeper sequencing was needed to define these groups. Notably, deep sequencing approaches continue to be used to validate and clarify the transcriptome of rare cell populations initially identified in 3' end counting datasets[24,78]. Our data further supports the value of increased scRNA-seq read depth and coverage, as well as focused analysis of specific cell types, to improve understanding of human photoreceptor development and its relationship to retinal disease.

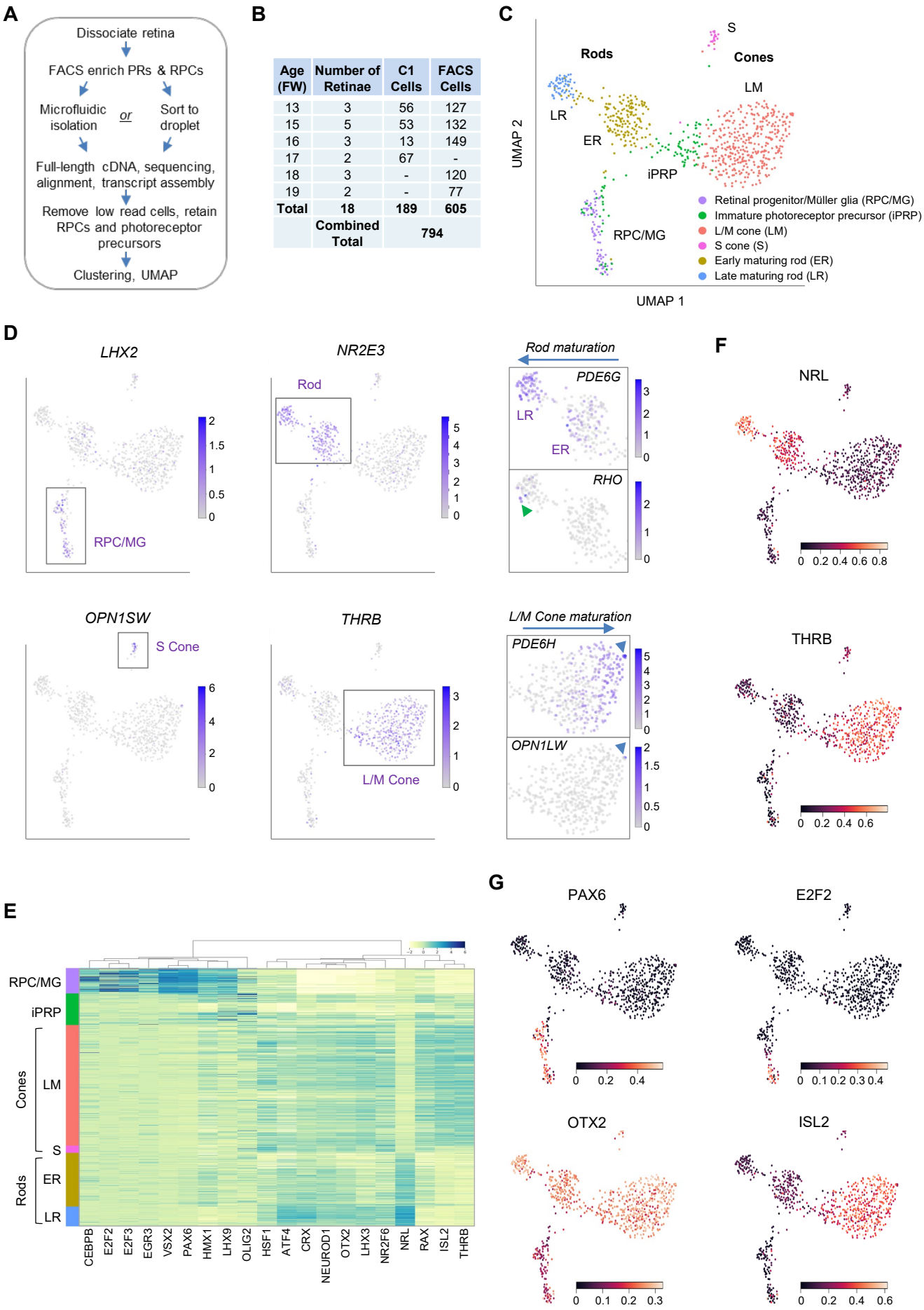
Acknowledgments

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. D.W.H.S was supported in part by a USC Provost Fellowship and a NIH T32 training grant (5T32HD060549) to the USC Development, Stem Cells, and Regenerative Medicine Department. This study was supported in part by an unrestricted grant to the USC Department of Ophthalmology from Research to Prevent Blindness, by NIH Center Core Grant P30CA014089 to USC Roski Eye Institute, and by the Larry and Celia Moh Foundation, the Neonatal Blindness Research Fund, the A. B. Reins Foundation, the Knights Templar Eye Foundation, and NIH grants R01CA137124 and R01EY026661 (to D.C.).

Author contributions: D.W.H.S., H.P.S., and D.C. designed initial experimental plans. D.W.H.S., S.L., J.B., J.G.A., Y. K, and H.P.S. performed experiments. K.S. and M.S. performed computational analyses. M.R. performed statistical analyses. M.T. and B.H.G. assisted with tissue collection. M.B and M.A.B. contributed bioinformatics approaches. C.M.C. provided a key reagent. D.W.H.S. and D.C. drafted the manuscript. All authors critiqued and approved the manuscript prior to submission.

Declaration of interests

The authors declare no competing interests.



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 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 from FACS or using C1 microfluidics. **B.** Retinae and cell numbers examined at each age. **C.** UMAP plot of final dataset, low resolution clusters labeled by cell type. **D.** 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*). **E.** Ward-clustered heatmaps of the highest scoring SCENIC regulons in each cluster, displaying Z-score normalized regulon activities. **F,G.** UMAP plots of regulon area under the curve (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.

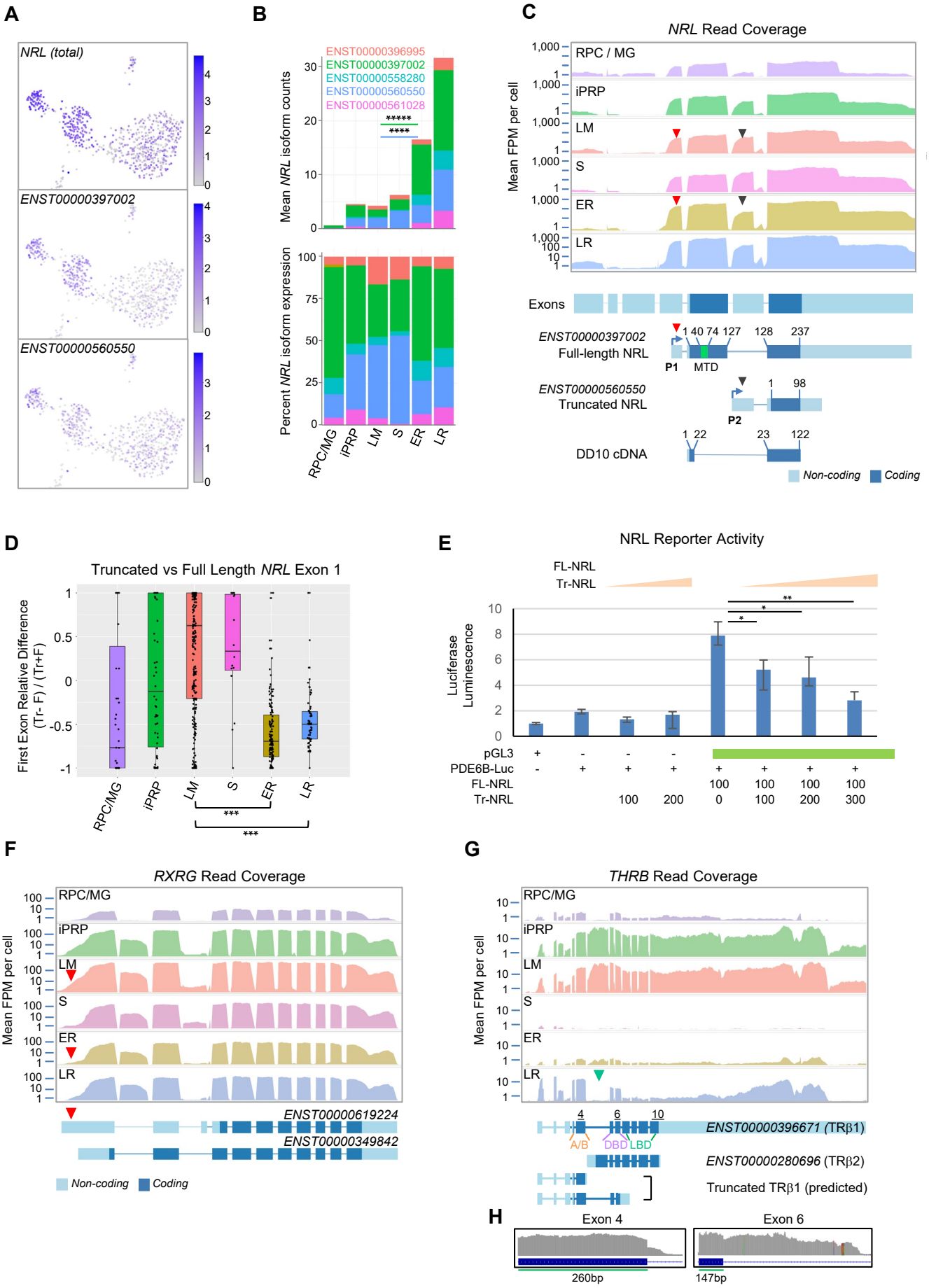

A
NRL (total)
ENST00000397002
ENST00000560550
B
ENST00000396995
ENST00000397002
ENST00000558280
ENST00000560550
ENST00000561028
Mean NRL isoform counts
Percent NRL isoform expression
RPC/MG
iPRP
LM
S
ER
LR
C
NRL Read Coverage
Mean FPM per cell
RPC / MG
iPRP
LM
S
ER
LR
Exons
ENST00000397002
Full-length NRL
P1
MTD
ENST00000560550
Truncated NRL
P2
DD10 cDNA
Non-coding
Coding
D
Truncated vs Full Length NRL Exon 1
First Exon Relative Difference
(Tr- F) / (Tr+F)
E
NRL Reporter Activity
FL-NRL
Tr-NRL
Luciferase
Luminescence
pGL3
PDE6B-Luc
F
RXRG Read Coverage
Mean FPM per cell
ENST00000619224
ENST00000349842
Non-coding
Coding
G
THRB Read Coverage
ENST00000396671 (TRβ1)
A/B
DBD
LBD
ENST00000280696 (TRβ2)
Truncated TRβ1 (predicted)
H
Exon 4
Exon 6
260bp
147bp

Figure 2: Alternative *NRL*, *RXRG*, and *THRB* 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.** Read coverage across Ensembl annotated *NRL* exons. *Top*: Mean read counts for cells in each cluster. *Bottom*: Transcript structures for *ENST00000397002, ENST00000560550,* and DD10, numbered according to amino acid positions. Minimal transactivation domain (MTD) in green. Arrowheads: First exons showing change in peak height (red = P1, black = P2). **D.** Relative difference 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.** *PDE6B*-luciferase reporter activity in NIH3T3 cells transfected with indicated amounts (ng) of pcDNA4C-EF1α FL-NRL and Tr-NRL vectors shown in Figure S4A. 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. **F, G.** Read coverage plot across annotated *RXRG* (F) and *THRB* (G) exons as in panel 2C. Arrowheads: Red = Extended *RXRG* 5' UTR in *ENST00000619224.* Green = First *THRB* exon in *ENST00000280696* encoding TRβ2. For *THRB,* exon numbers[48] are indicated above and protein domains (A/B, DNA-binding (DBD), and ligand binding (LBD)) below TRβ1. **H.** Compiled read coverage for LR cells across *THRB* exons 4 and 6 and expressed intron 4 and 6 sequences. See also Figure S2.

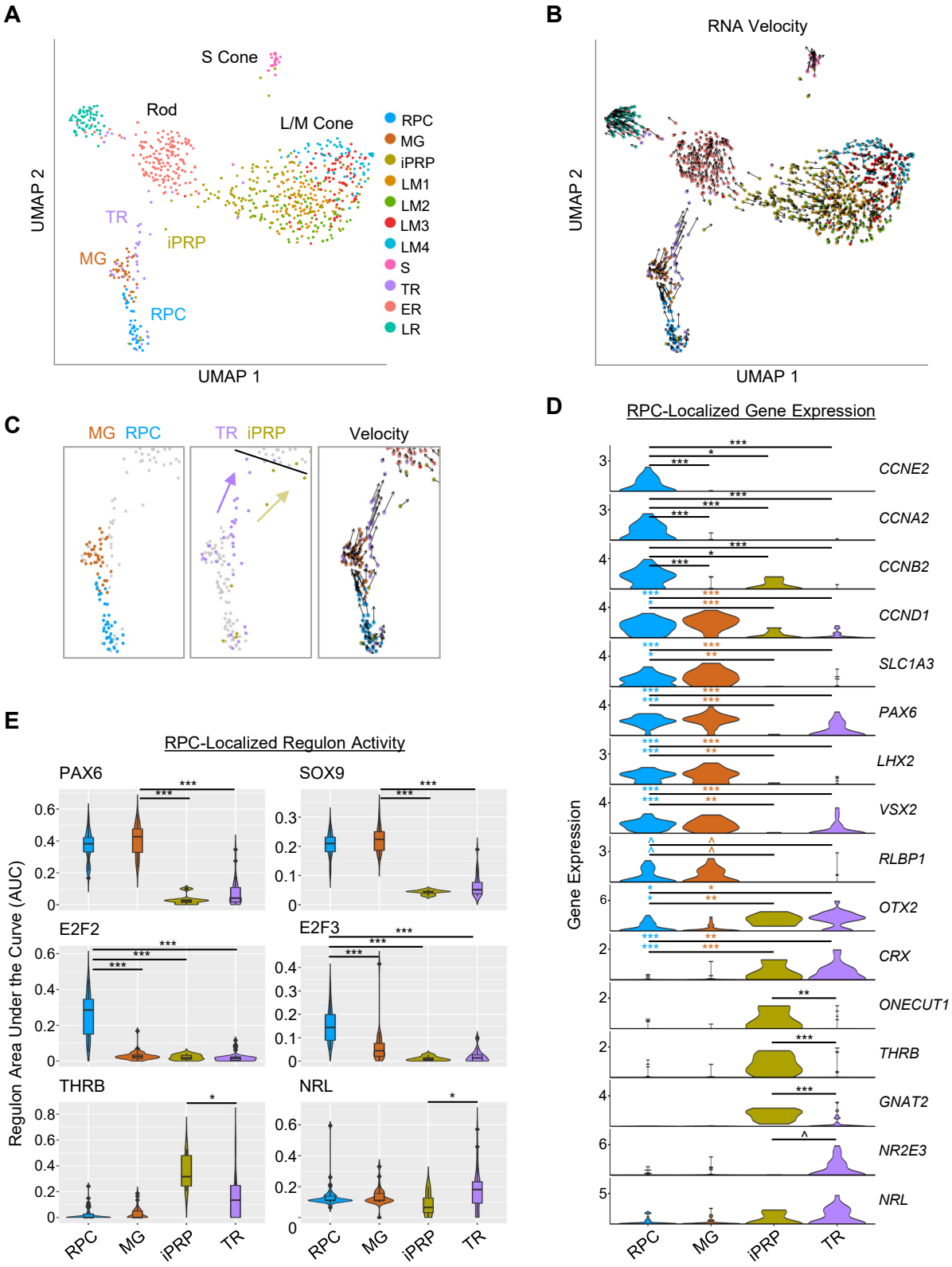

A
S Cone
Rod
L/M Cone
TR
iPRP
MG
RPC
UMAP 2
UMAP 1
RPC
MG
iPRP
LM1
LM2
LM3
LM4
S
TR
ER
LR
B
RNA Velocity
UMAP 2
UMAP 1
C
MG RPC
TR iPRP
Velocity
D
RPC-Localized Gene Expression
Gene Expression
CCNE2
CCNA2
CCNB2
CCND1
SLC1A3
PAX6
LHX2
VSX2
RLBP1
OTX2
CRX
ONECUT1
THRB
GNAT2
NR2E3
NRL
RPC
MG
iPRP
TR
E
RPC-Localized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

Figure 3: 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). See also Figure S3.

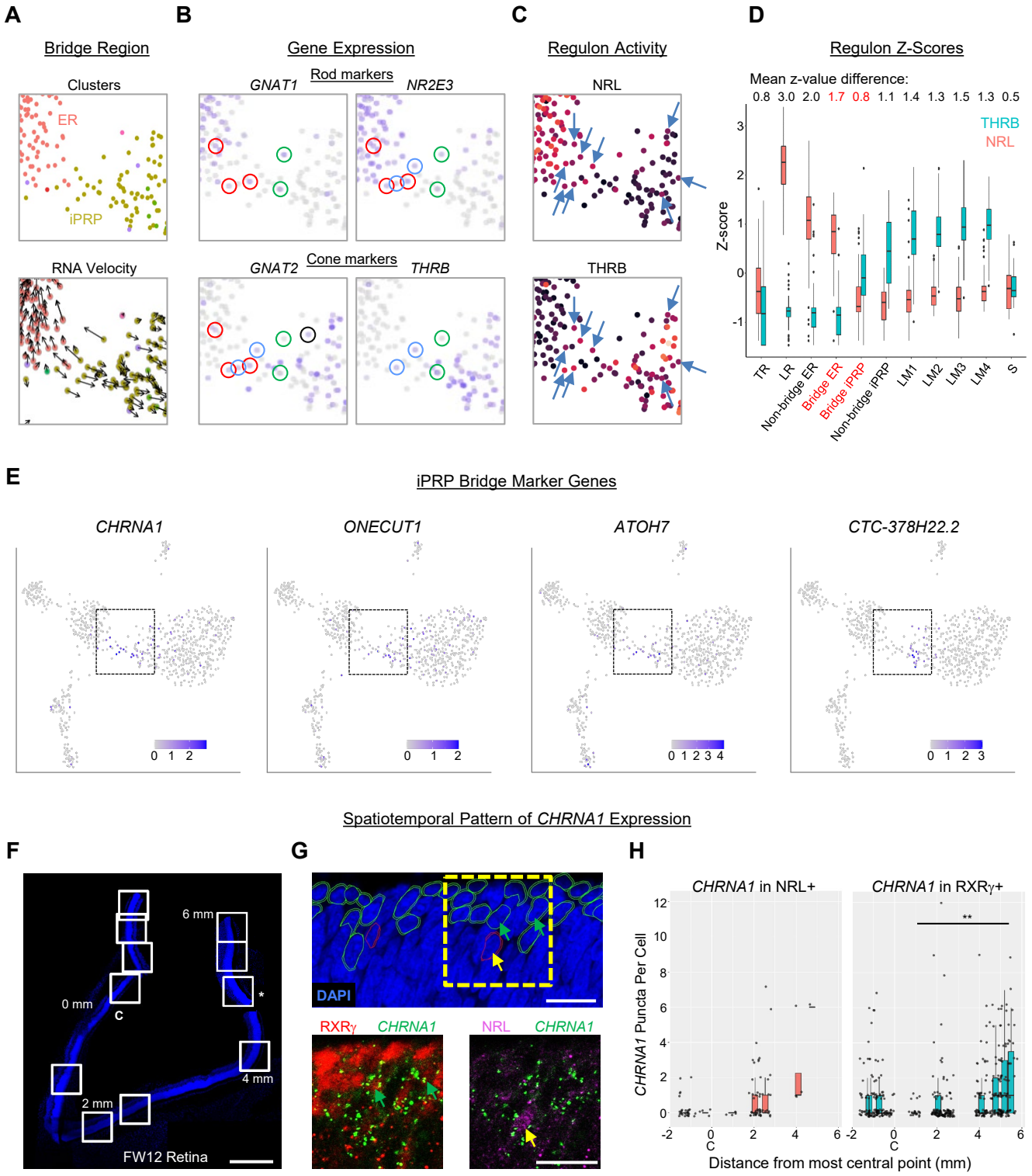

A
Bridge Region
Clusters
ER
iPRP
RNA Velocity
B
Gene Expression
Rod markers
GNAT1
NR2E3
Cone markers
GNAT2
THRB
C
Regulon Activity
NRL
THRB
D
Regulon Z-Scores
Mean z-value difference:
0.8 3.0 2.0 1.7 0.8 1.1 1.4 1.3 1.5 1.3 0.5
THRB
NRL
Z-score
TR
LR
Non-bridge ER
Bridge ER
Bridge iPRP
Non-bridge iPRP
LM1
LM2
LM3
LM4
S
E
iPRP Bridge Marker Genes
CHRNA1
ONECUT1
ATOH7
CTC-378H22.2
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 4: **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 with StarDist 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). See also Figure S4.

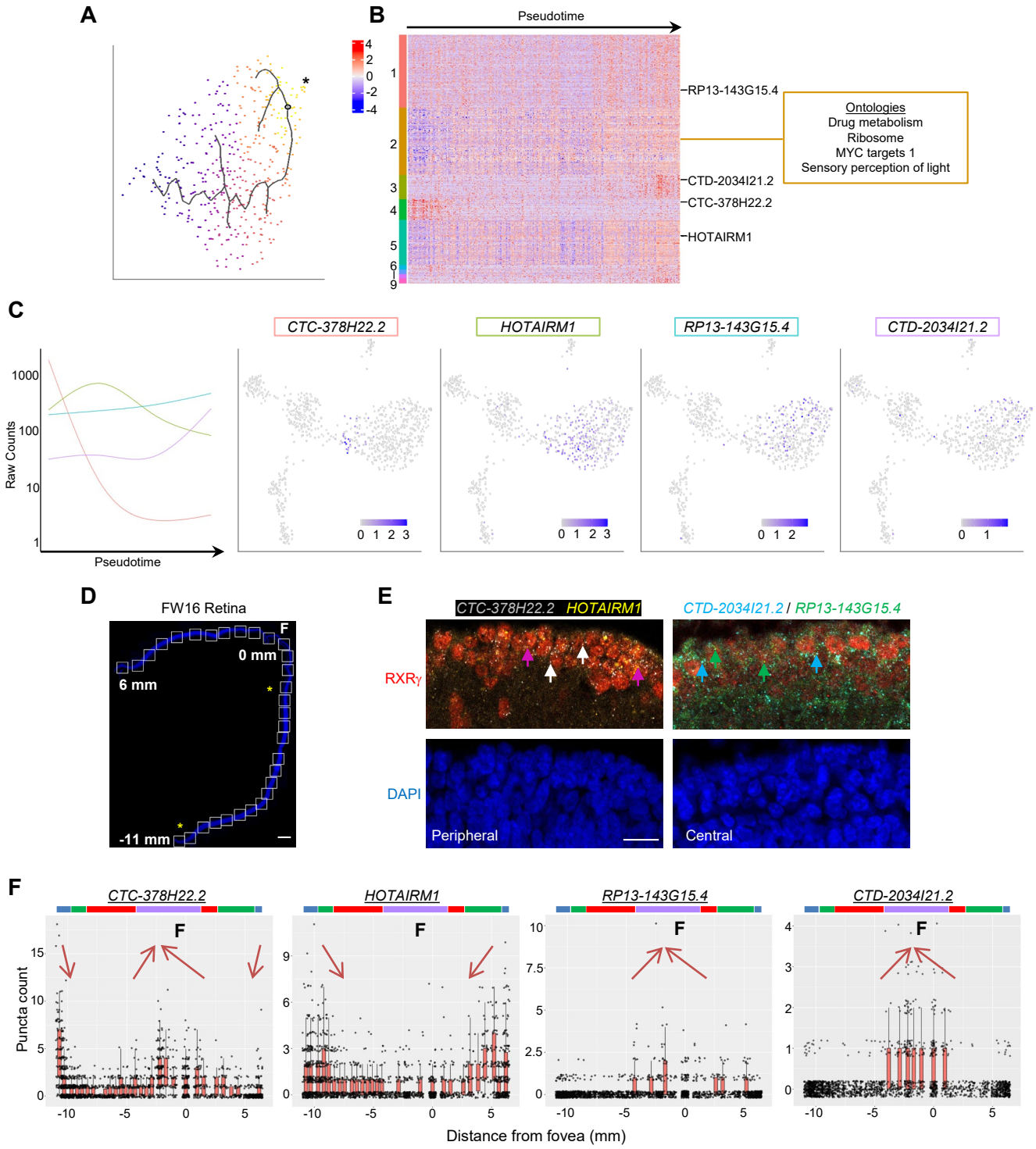

A
B
Pseudotime
RP13-143G15.4
CTD-2034I21.2
CTC-378H22.2
HOTAIRM1
Ontologies
Drug metabolism
Ribosome
MYC targets 1
Sensory perception of light
C
CTC-378H22.2
HOTAIRM1
RP13-143G15.4
CTD-2034I21.2
Raw Counts
Pseudotime
D
FW16 Retina
0 mm
6 mm
-11 mm
E
CTC-378H22.2 HOTAIRM1
CTD-2034I21.2 / RP13-143G15.4
RXRγ
DAPI
Peripheral
Central
F
CTC-378H22.2
HOTAIRM1
RP13-143G15.4
CTD-2034I21.2
Puncta count
Distance from fovea (mm)

Figure 5: An early L/M cone pseudotemporal trajectory with successive lncRNA gene expression. A. Pseudotemporal trajectory through the L/M cone population derived with Monocle 3. *: Root cell used to define endpoint. **B.** Heatmaps of genes in Monocle defined modules 1-9, highlighting lncRNAs of interest and module 2 ontologies. **C.** Trendlines of raw count expression (*left*) and UMAP plots for lncRNAs correlating with early or late-upregulating modules. Line color matched to label boxes. **D,E.** Combined RXRγ immunohistochemical staining and FISH of lncRNAs on FW16 retina. **D**. 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 **E**. 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. **F.** Quantitation of lncRNA fluorescent puncta assigned to RXRγ+ cells after segmentation. Colored bars mark lncRNA expression regions as described in the text. See also Figure S5. Arrows indicate significant count differences between images ($p < 0.05$, Wald test).

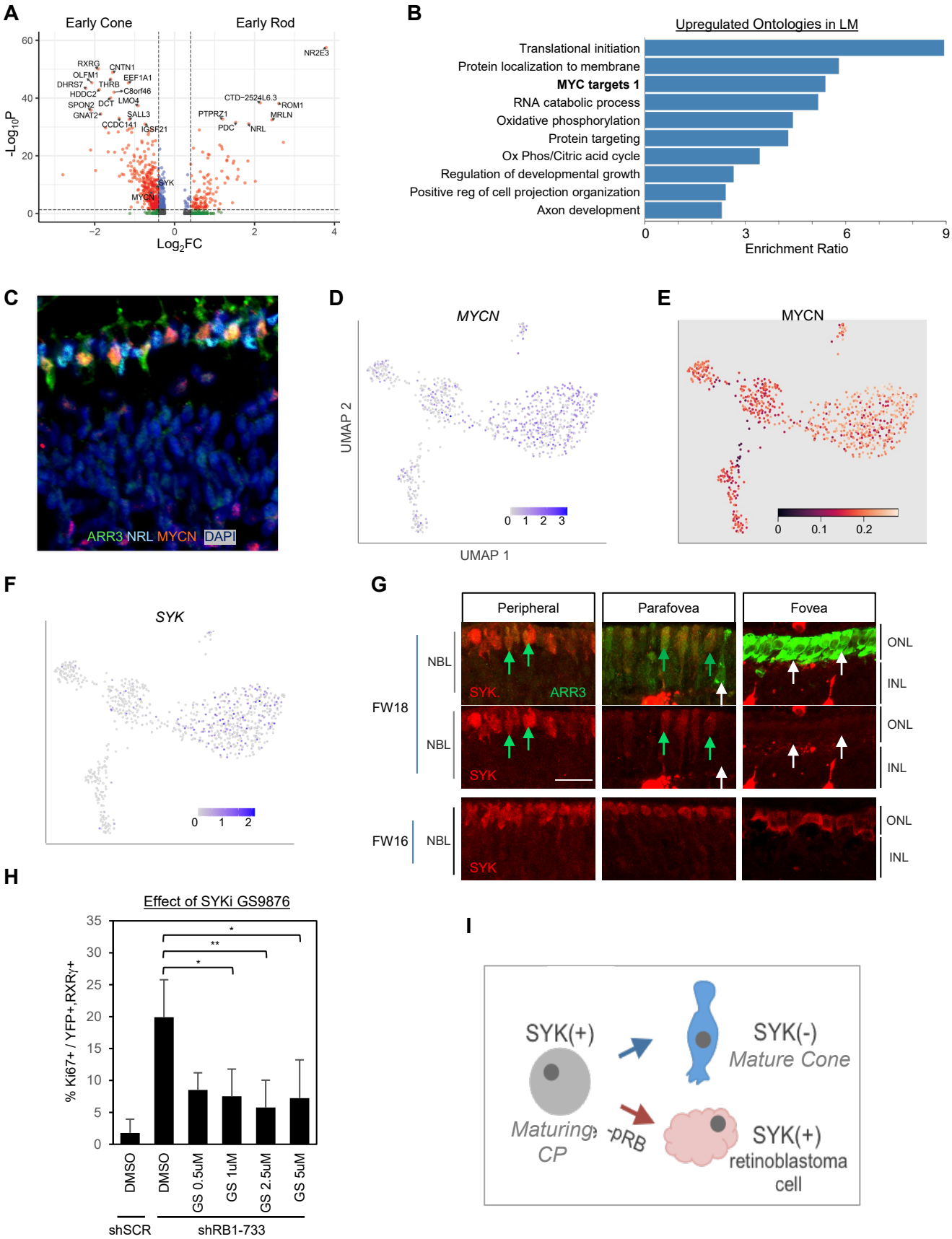

A
Early Cone
Early Rod
-Log10P
Log2FC
RXRG
CNTN1
OLFM1
EEF1A1
DHRS7
THRB
HDDC2
C8orf46
SPON2
DCT
LMO4
GNAT2
SALL3
CCDC141
IGSF21
SYK
MYCN
NR2E3
CTD-2524L6.3
ROM1
PTPRZ1
MRLN
PDC
NRL
B
Upregulated Ontologies in LM
Translational initiation
Protein localization to membrane
MYC targets 1
RNA catabolic process
Oxidative phosphorylation
Protein targeting
Ox Phos/Citric acid cycle
Regulation of developmental growth
Positive reg of cell projection organization
Axon development
Enrichment Ratio
C
ARR3 NRL MYCN DAPI
D
MYCN
UMAP 2
UMAP 1
E
MYCN
F
SYK
G
Peripheral
Parafovea
Fovea
FW18
FW16
NBL
ONL
INL
SYK
ARR3
H
Effect of SYKi GS9876
% Ki67+ / YFP+,RXRγ+
DMSO
GS 0.5uM
GS 1uM
GS 2.5uM
GS 5uM
shSCR
shRB1-733
I
SYK(+)
Maturing CP
pRB
SYK(-)
Mature Cone
SYK(+)
retinoblastoma cell

Figure 6: 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. Labeled genes: pAdj <10e-32, except for *GNAT1, PDE6H, SYK, MYCN.* pAdj cutoff =0.05, $\log_2$FC cutoff = |0.4|. **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, E.** UMAP plots of *MYCN* expression (D) and regulon activity (E)**. F.** *SYK* expression UMAP plot. **G.** Immunohistochemical analysis of SYK expression in FW18 and FW16 retinae with rabbit polyclonal antibody 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.

STAR Methods

Resource Availability

Lead contact: Further information and requests for resources should be directed to the Lead Contact, David Cobrinik (dcobrinik@chla.usc.edu).

Materials availability: DNA constructs are available upon request.

Data and code availability: 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, SCENIC regulon activity, and sample metadata; cluster marker gene and violin plots; exon coverage plots; Ensembl isoform assignments) is pending. Single-cell RNA-sequencing files and processed final gene and transcript matrices are publicly available and have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus under accession number GSE207902.

Experimental Model and Subject Details

Primary human tissue

Human fetal samples were 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[79]. Each retina subjected to sequencing was from a different donor.

Method Details

Retinal dissociation and RNA-sequencing

Single cell isolation

Retinae were dissected and dissociated in ice-cold phosphate-buffered saline (PBS) as described[13]. 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[13] 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 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 in the morning 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). Seq 1 retinae were sequenced on the Illumina NextSeq 500 (2x75) and all others on Illumina HiSeq 4000 (2x75).

<u>Computational Methods</u>

Read processing and dimensionality reduction

FASTQ files had adapter sequences removed using the *trimgalore* wrapper for *Cutadapt* (Martin 2011). Trimmed FASTQ files were used as input for *HISAT2*[80] 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*[81], 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[82] workflow accessible at https://github.com/whtns/ARMOR.

Dimensionality reduction and data visualization were performed using the Seurat toolset[83]. 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[84,85]. 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)[86].

Read coverage plots for genes of interest were generated using *wiggleplotr* to visualize BigWig files with ENSEMBL isoforms. Individual exon counts were identified using *DEXseq*[87], 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 (Figure 2D). Reads crossing target splice donor sites were evaluated as spliced or unspliced from BAM files to determine intron readthrough.

Differential expression analysis

To identify marker features for each cluster, Wilcoxon-rank sum tests and specificity scores were computed using the *genesorteR* 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*.

Overrepresentation analyses were performed with WebGestalt[88]. 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)[39,89]. 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*[49]. 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[60]. 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 were grouped into modules of co-regulation by performing UMAP and Louvain community analysis. This gene list was further filtered before analysis by excluding genes without expression > 0.5 in at least 5% of the cells present in the pseudotime (Table S5).

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-immunostaining, 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, were mixed in blocking solution and applied to samples overnight (~16-18 hrs) 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 immunostaining of tissue sections and cells on coverslips was performed as described[14] but without initial EDTA washing steps.

In situ hybridization and Immunostaining

Human retinal sections were prepared as above but using RNase-free reagents for all stages (PBS, TBS) and ultrapure sucrose (JT Baker, 4097-04). Cryosectioning instruments were cleaned with a mixture of ELIMINase (Decon Labs, Inc) and 70% ethanol prior to use.

RNA probes were designed by and ordered from Molecular Instruments, Inc, with a goal of 20x20 bp probes per target RNA sequence. The accession numbers for target sequences are shown in Key Resources table, 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.

Quantitation of FISH

For *in situ* hybridization, QuPath[90] 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[91] 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. Remaining annotations were classified using single object classifiers such as RXRγ+, NRL+, or GFP+/- as needed. 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 = 2 um^2) 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.

Lentivirus production

Four plasmids were each used in the production of three different lentiviral constructs (pMDL, pVSVG, pREV and a lentiviral vector)[92]. 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 polyethylenimine (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 hrs 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 hrs, 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[46] 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 hrs to a final

concentration of 10 µM and cells incubated for an additional 4 hrs, 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[93] (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*[94] 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 hrs, 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 hrs 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[14], 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.

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

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*)[13] 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 hrs, fixed in 4% PFA for 10 min, washed in 1x PBS three times, and stored at -20°C until immunostaining.

Statistical Analysis

Statistical methods and packages are cited in the STAR 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.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat polyclonal anti-NRL (WB: 1:2000-4000)	R&D Systems	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
Rabbit polyclonal anti-RXRγ (1:800)	Santa Cruz	CAT# SC-555, RRID: AB_2269865
Mouse monoclonal anti-KI67 (1:200)	BD Bioscience	CAT# 550609, RRID: AB_393778
Goat polyclonal anti-GFP (1:500)	Abcam	CAT# ab6673, RRID: AB_305643
Mouse monoclonal anti-CD133-PE (1:50)	Miltenyi Biotec	CAT# 130-113-108, RRID: AB_2725937
Mouse monoclonal anti-CD44-FITC (1:50)	BD Biosciences	CAT# 555478, RRID: AB_395870
Mouse monoclonal anti-CD49b-FITC (1:10)	BD Biosciences	CAT# 555498, RRID: AB_395888
Donkey polyclonal anti-goat IgG Alexa Fluor 488 (1:300)	Jackson Laboratories	CAT# 705-545-147, RRID: AB_2336933
Donkey polyclonal anti-mouse IgG Alexa Fluor 680 (1:500)	Life Technologies	CAT# A10038, RRID: AB_2534014
Donkey polyclonal anti-mouse IgG Alexa Fluor 680 (1:500)	Invitrogen	CAT# A11057, RRID: AB_2534104
Biological samples		
Human fetal retina	Family Planning Associates, Los Angeles, CA	Not applicable
Human fetal retina	Advanced Bioscience Resources, Alameda, CA	Not applicable
Chemicals, peptides, and recombinant proteins		
Molecular Instruments Probe – *CHRNA1*	Molecular Instruments Inc.	Target sequence NCBI accession number: NM_000079.4
Molecular Instruments Probe – *RP13-143G15.4*	Molecular Instruments Inc.	
Molecular Instruments Probe – *CTC-378H22.2*	Molecular Instruments Inc.	Target sequence ENSEMBL ID: ENST00000559786.1
Molecular Instruments Probe – *HOTAIRM1*	Molecular Instruments Inc.	Target sequence NCBI accession number: NR_038366.1

Molecular Instruments Probe – *CTD-2034I21.2*	Molecular Instruments Inc.	Target sequence NCBI accession number: XR_001752169.1
GS-9876 SYK inhibitor	MedChemExpress	CAT# HY-109091
Critical commercial assays		
SMARTer Ultra Low RNA Kit for the Fluidigm C1 System	Clontech (Takara Bio)	CAT# 634835/634935
SMART-Seq V4 Ultra Low Input RNA Kit	Takara Bio	CAT# 63891
Nextera XT DNA Library Preparation Kit	Illumina	CAT# FC-131-1096
Nano-Glo Dual-Luciferase Reporter Assay System	Promega, Inc	CAT# N1610
Deposited data		
Single-cell RNA sequencing	This paper	GEO: GSE207802
Experimental models: Cell lines		
CHLA-VC-RB-31	Stachelek et al., 2022	N/A
HEK-293T	ATCC	N/A
NIH-3T3	ATCC	N/A
Oligonucleotides		
1) Del-His F	N/A	CCGAAACCATGGCCCTGCCCCCCAGC
2) Del-His R	N/A	GGGCCATGGTTTCGGAGGCCGTCCG
3) NRL-no-His F	N/A	CCGAAACCATGTCTGTGCGGGAGCTAAACC
4) NRL-no-His R	N/A	CAGACATGGTTTCGGAGGCCGTCCG
5) pcDNA del NRL F	N/A	CCGAAACCGCCGTTCAGAGCACCTTGTGG
6) pcDNA del NRL R	N/A	GAACGGCGGTTTCGGAGGCCGTCCG
7) PDE -93 F IF	N/A	TCTTACGCGTGCTAGAGCGCAGGCCCCCATTTG
8) PDE +53 R IF	N/A	CTTAGATCGCAGATCGGTGGCTGCCTGTCCCTG
9) pGL3-sv40 SDM F	N/A	CTGCGATCAAGCTTGGCATTCCGGTACTG
10) pGL3-sv40 SDM R	N/A	CAAGCTTGATCGCAGATCGGTGGCTG
11) pGL3 del SV40 R	N/A	CAAGCTTGATCGCAGATCTCGAGCCC
12) FL-NRL IF pU-G F	N/A	GCCTTCTAGAGGATCCATGTCTGTGCGGGAGCTAAACC
13) FL-NRL IF pU-G R	N/A	CGCCGGAGCCGGATCCTCAGAGGAAGAGGTGGGAGGG
Recombinant DNA		
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 and algorithms		
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	http://ccb.jhu.edu/software/stringtie/
Snakemake/ARMOR	Orjuela et al., 2019	https://github.com/whtns/ARMOR
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. Young, R.W. (1985). Cell differentiation in the retina of the mouse. Anat Rec *212*, 199–205. 10.1002/ar.1092120215.

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

3. Miesfeld, J.B., and Brown, N.L. (2019). Eye organogenesis: A hierarchical view of ocular development. In Current Topics in Developmental Biology (Elsevier Inc.), pp. 351–393. 10.1016/bs.ctdb.2018.12.008.

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

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

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

7. Sehgal, R., Andres, D.J., Adler, R., and Belecky-Adams, T.L. (2006). Bone Morphogenetic Protein 7 Increases Chick Photoreceptor Outer Segment Initiation. Investigative Opthalmology & Visual Science *47*, 3625. 10.1167/iovs.06-0079.

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

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

10. Hussey, K.A., Hadyniak, S.E., and Johnston, R.J. (2022). Patterning and Development of Photoreceptors in the Human Retina. Front Cell Dev Biol *10*, 1–21. 10.3389/fcell.2022.878350.

11. Seabrook, T.A., Burbridge, T.J., Crair, M.C., and Huberman, A.D. (2017). Architecture, Function, and Assembly of the Mouse Visual System. Annu Rev Neurosci *40*, 499–538. 10.1146/annurev-neuro-071714-033842.

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

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

14. Singh, H.P., Wang, S., Stachelek, K., Lee, S., Reid, M.W., Thornton, M.E., Craft, C.M., Grubbs, B.H., and Cobrinik, D. (2018). Developmental stage-specific proliferation and retinoblastoma genesis in RB-deficient human but not mouse cone precursors. Proceedings of the National Academy of Sciences *115*, E9391–E9400. 10.1073/pnas.1808903115.

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

16. Kim, J.-W., Yang, H.-J., Oel, A.P., Brooks, M.J., Jia, L., Plachetzki, D.C., Li, W., Allison, W.T., and Swaroop, A. (2016). Recruitment of Rod Photoreceptors from Short-Wavelength-Sensitive Cones during the Evolution of Nocturnal Vision in Mammals. Dev Cell *37*, 520–532. 10.1016/j.devcel.2016.05.023.

17. Kozulin, P., and Provis, J.M. (2009). Differential gene expression in the developing human macula: Microarray analysis using rare tissue samples. J Ocul Biol Dis Infor *2*, 176–189. 10.1007/s12177-009-9039-1.

18. Hoshino, A., Horvath, S., Sridhar, A., Chitsazan, A., and Reh, T.A. (2019). Synchrony and asynchrony between an epigenetic clock and developmental timing. Sci Rep *9*, 3770. 10.1038/s41598-019-39919-3.

19. Aldiri, I., Xu, B., Wang, L., Chen, X., Hiler, D., Griffiths, L., Valentine, M., Shirinifard, A., Thiagarajan, S., Sablauer, A., et al. (2017). The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis. Neuron *94*, 550-568.e10. 10.1016/j.neuron.2017.04.022.

20. Mellough, C.B., Bauer, R., Collin, J., Dorgau, B., Zerti, D., Dolan, D.W.P., Jones, C.M., Izuogu, O.G., Yu, M., Hallam, D., et al. (2019). An integrated transcriptional analysis of the developing human retina. Development *146*. 10.1242/dev.169474.

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

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

23. Sridhar, A., Hoshino, A., Finkbeiner, C.R., Chitsazan, A., Dai, L., Haugan, A.K., Eschenbacher, K.M., Jackson, D.L., Trapnell, C., Bermingham-McDonogh, O., et al. (2020). Single-Cell Transcriptomic Comparison of Human Fetal Retina, hPSC-Derived

Retinal Organoids, and Long-Term Retinal Cultures. Cell Rep *30*, 1644-1659.e4. 10.1016/j.celrep.2020.01.007.

24. lo Giudice, Q., Leleu, M., la Manno, G., and Fabre, P.J. (2019). Single-cell transcriptional logic of cell-fate specification and axon guidance in early born retinal neurons. Development *146*. 10.1242/dev.178103.

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

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

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

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

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

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

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

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

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

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

35. Daum, J.M., Keles, Ö., Holwerda, S.J.B., Kohler, H., Rijli, F.M., Stadler, M., and Roska, B. (2017). The formation of the light-sensing compartment of cone photoreceptors coincides with a transcriptional switch. Elife *6*, e31437. 10.7554/eLife.31437.

36. O'Brien, K.M.B., Schulte, D., and Hendrickson, A.E. (2003). Expression of photoreceptor-associated molecules during human fetal eye development. Mol Vis *9*, 401–409.

37. Voigt, A.P., Mullin, N.K., Stone, E.M., Tucker, B.A., Scheetz, T.E., and Mullins, R.F. (2021). Single-cell RNA sequencing in vision research: Insights into human retinal health and disease. Prog Retin Eye Res *83*, 100934. 10.1016/j.preteyeres.2020.100934.

38. Xiao, M., and Hendrickson, A. (2000). Spatial and temporal expression of short, long/medium, or both opsins in human fetal cones. J Comp Neurol *425*, 545–559. 10.1002/1096-9861(20001002)425:4<545::AID-CNE6>3.0.CO;2-3.

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

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

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

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

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

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

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

46. Stachelek, K., Harutyunyan, N., Lee, S., Beck, A., Kim, J., Berry, J.L., Nagiel, A., Reynolds, C.P., Murphree, A.L., Lee, T.C., et al. (2022). Non-Synonymous, Synonymous, and Non-Coding Nucleotide Variants Contribute to Recurrently Altered Biological

Processes During Retinoblastoma Progression. Preprint at bioRxiv.10.27.512289. 10.1101/2022.10.27.512289.

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

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

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

50. Roesch, K., Jadhav, A.P., Trimarchi, J.M., Stadler, M.B., Roska, B., Sun, B.B., and Cepko, C.L. (2008). The transcriptome of retinal Müller glial cells. J Comp Neurol *509*, 225–238. 10.1002/cne.21730.

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

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

53. Emerson, M.M., Surzenko, N., Goetz, J.J., Trimarchi, J., and Cepko, C.L. (2013). Otx2 and Onecut1 Promote the Fates of Cone Photoreceptors and Horizontal Cells and Repress Rod Photoreceptors. Dev Cell *26*, 59–72. 10.1016/j.devcel.2013.06.005.

54. Finkbeiner, C., Ortuño-Lizarán, I., Sridhar, A., Hooper, M., Petter, S., and Reh, T.A. (2022). Single-cell ATAC-seq of fetal human retina and stem-cell-derived retinal organoids shows changing chromatin landscapes during cell fate acquisition. Cell Rep *38*, 110294. 10.1016/j.celrep.2021.110294.

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

56. Poché, R.A., Furuta, Y., Chaboissier, M.-C., Schedl, A., and Behringer, R.R. (2008). Sox9 is expressed in mouse multipotent retinal progenitor cells and functions in Müller Glial cell development. J Comp Neurol *510*, 237–250. 10.1002/cne.21746.

57. Feng, L., Xie, Z., Ding, Q., Xie, X., Libby, R.T., and Gan, L. (2010). MATH5 controls the acquisition of multiple retinal cell fates. Mol Brain *3*, 36. 10.1186/1756-6606-3-36.

58. Zhang, X.-M., Hashimoto, T., Tang, R., and Yang, X.-J. (2018). Elevated expression of human bHLH factor ATOH7 accelerates cell cycle progression of progenitors and enhances production of avian retinal ganglion cells. Sci Rep *8*, 6823. 10.1038/s41598-018-25188-z.

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

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

61. Qu, C., Zheng, D., Li, S., Liu, Y., Lidofsky, A., Holmes, J.A., Chen, J., He, L., Wei, L., Liao, Y., et al. (2018). Tyrosine kinase SYK is a potential therapeutic target for liver fibrosis. Hepatology *68*, 1125–1139. 10.1002/hep.29881.

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

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

64. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214. 10.1016/j.cell.2015.05.002.

65. Liang, Q., Dharmat, R., Owen, L., Shakoor, A., Li, Y., Kim, S., Vitale, A., Kim, I., Morgan, D., Liang, S., et al. (2019). Single-nuclei RNA-seq on human retinal tissue provides improved transcriptome profiling. Nat Commun *10*, 5743. 10.1038/s41467-019-12917-9.

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

67. Oh, E.C.T., Cheng, H., Hao, H., Jia, L., Khan, N.W., and Swaroop, A. (2008). Rod differentiation factor NRL activates the expression of nuclear receptor NR2E3 to suppress the development of cone photoreceptors. Brain Res *1236*, 16–29. 10.1016/j.brainres.2008.01.028.

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

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

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

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

72. Mizeracka, K., DeMaso, C.R., and Cepko, C.L. (2013). Notch1 is required in newly postmitotic cells to inhibit the rod photoreceptor fate. Development *140*, 3188–3197. 10.1242/dev.090696.

73. Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science *360*, 981–987. 10.1126/science.aar4362.

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

75. Rushlow, D.E., Mol, B.M., Kennett, J.Y., Yee, S., Pajovic, S., Thériault, B.L., Prigoda-Lee, N.L., Spencer, C., Dimaras, H., Corson, T.W., et al. (2013). Characterisation of retinoblastomas without RB1 mutations: genomic, gene expression, and clinical studies. Lancet Oncol *14*, 327–334. 10.1016/S1470-2045(13)70045-7.

76. Singh, H.P., Shayler, D.W.H., Fernandez, G.E., Thornton, M.E., Craft, C.M., Grubbs, B.H., and Cobrinik, D. (2022). An immature, dedifferentiated, and lineage-deconstrained cone precursor origin of N-Myc–initiated retinoblastoma. Proceedings of the National Academy of Sciences *119*, e2200721119. 10.1073/pnas.2200721119.

77. Liu, H., Zhang, Y., Zhang, Y.Y., Li, Y.P., Hua, Z.Q., Zhang, C.J., Wu, K.C., Yu, F., Zhang, Y., Su, J., et al. (2020). Human embryonic stem cell-derived organoid retinoblastoma reveals a cancerous origin. Proc Natl Acad Sci U S A *117*, 33628–33638. 10.1073/PNAS.2011780117.

78. Davie, K., Janssens, J., Koldere, D., de Waegeneer, M., Pech, U., Kreft, Ł., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., et al. (2018). A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell *174*, 982-998.e20. 10.1016/j.cell.2018.05.057.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

94. Lerner, L.E., Gribanova, Y.E., Ji, M., Knox, B.E., and Farber, D.B. (2001). Nrl and Sp Nuclear Proteins Mediate Transcription of Rod-specific cGMP-phosphodiesterase β-

Subunit Gene. Journal of Biological Chemistry *276*, 34999–35007. 10.1074/jbc.M103301200.