UNIVERSITY OF CALIFORNIA SAN DIEGO

Validation of Human Retinal Organoid Derived Photoreceptors through Single-Cell RNA Sequencing

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Manan Chopra

Committee in charge:

Professor Karl Jonas Wahlin, Chair
Professor Michael William Perry, Co-Chair
Professor Deborah Yelon

2024

The thesis of Manan Chopra is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego
2024

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Mom, Dad, Riya and the rest of my extended family and friends for being there for me for as long as I can remember. Their constant support through grade school, undergraduate and graduate school is a big reason I am in this position today and played a large role in making me who I am. They mean the world to me.

I would like to acknowledge Aleksandra Dragozova and Grace Kim for their unending support (both professionally and otherwise), and invaluable help in reviewing and writing this thesis. I am forever thankful for their friendship, and words cannot express how grateful I am to have them in my life.

I would like to acknowledge Devansh Agarwal for being an incredible friend and mentor to me, answering question after question (there were many through the course of my time at the lab) and always looking to help. I would additionally like to acknowledge Devansh for help with the library preparations carried out to generate data for this thesis.

I would like to acknowledge Nate Kemmerer for prior work relating to the retinal organoids I analyze in this thesis, and for (usually) laughing at my jokes.

I would like to acknowledge Sonia Rivero, Kavitha Thirumaran and the rest of my peers at the Wahlin Lab for their constant support and help throughout my three years with them. I could not have asked for a better group of people to do research with.

I would like to acknowledge Dr. Michael Perry and Dr. Deborah Yelon for serving on my committee and taking the time to review my manuscript and attend my defense.

Lastly, I would like to acknowledge my mentor Dr. Karl Wahlin for giving me the incredible opportunity to work in this field, and for his constant support and guidance throughout my journey.

VITA

2023	Bachelor of Science, University of California San Diego
2021-2024	Research Assistant, University of California San Diego
2023-2024	Teaching Assistant, University of California San Diego
2024	Master of Science, University of California San Diego

PUBLICATIONS

"Accessibility and Transcriptional Differences in Human Stem Cell-Derived Early-Stage Retinal Organoids" *Cells* 11, 3412 (2022).

"Human retinal ganglion cell neurons generated by synchronous BMP inhibition and transcription factor mediated reprogramming" *npj Regen Med* 8, 1–18 (2023).

"Restoring vision and rebuilding the retina by Müller glial cell reprogramming" *Stem Cell Res* 66, 103006 (2023).

"Early Transcriptional Dynamics of Retinal Ganglion Cells Generated by Direct Conversion" *iovs* 64, (2023).

FIELDS OF STUDY

Major Field: Biology

Studies in Bioinformatics
Studies in Molecular Biology

ABSTRACT OF THE THESIS

Validation of Human Retinal Organoid Derived Photoreceptors through Single-Cell RNA Sequencing

by

Manan Chopra

Master of Science in Biology

University of California San Diego, 2024

Professor Karl Jonas Wahlin, Chair
Professor Michael William Perry, Co-Chair

Retinal degeneration (RD) is a leading cause of vision impairment and blindness worldwide. Diseases which cause RD, including glaucoma, retinitis pigmentosa (RP) and age-related macular degeneration (AMD) are characterized by dysfunction and/or loss of key retinal neurons. Studying these human diseases in model organisms is useful but often approaches a limit at which the model organism does not sufficiently capture the human context. Many groups are utilizing *in vitro* methods to study the human retina, the most prominent being human induced-pluripotent stem cell (iPSC) derived neurons and iPSC-derived 3D retinal organoids (rORGs). A key first step to studying diseases like AMD and

RP is establishing the accuracy of these kinds of *in vitro* models that can recapitulate the complexity of the retina and its many cell types. To address this need, I explored transcriptomic patterns in rORGs focusing on photoreceptors (PRs). I found that the rORG-derived PRs (O-PRs) exhibited a PR-like identity with both rod and cone markers detected. Importantly, I also compared these *in vitro* O-PRs with *in vivo* datasets and found that they share similarities in their transcriptomic profiles. This work is intended to serve as a resource for scientists seeking to perform similar analyses on retinal datasets as well as cell types beyond the nervous system. Toward this goal, I have documented all the code used to generate results in a publicly available repository with guides on how to perform the analysis regardless of prior bioinformatics experience.

Introduction

Retinal degenerative diseases often result in vision impairment or complete blindness, conditions which are estimated to affect almost 285 million adults worldwide.[1] Retinitis pigmentosa (RP) and age-related macular degeneration (AMD) are among the most prevalent causes of blindness and are both characterized by dysfunction of key retinal neurons.[2] This has led to a large effort by many groups to explore ways to rescue or regenerate lost neurons. If successful, this may reverse retinal degenerations (RDs) caused by a disruption of the intricate interplay between neurons and other cells in the retina.

The retina is a highly organized, layered tissue that lines the inside of the eye. Through a process known as phototransduction, light is converted by the retina into electro-chemical signals that filter through different retinal cell types and are sent to the brain. Briefly, light enters the eye through the cornea and is first detected by the photoreceptors (PRs), which are supplied with nutrients and support by the retinal pigment epithelium (RPE).[3] The PRs then propagate electrical signals to bipolar cells, which synapse with retinal ganglion cells (RGCs). RGC axons bundle together to form the optic nerve, which exits the back of the eye and sends electrical signals to the brain. Amacrine and horizontal cells serve to refine vision through lateral inhibition and/or neuromodulation, and Müller glia (MG) provide support and maintain retinal homeostasis (Fig 1a).[4] During retinal development, these retinal cell types arise in overlapping waves, broadly classified into early and late retinogenesis. RGCs, horizontal cells, amacrine cells and cone PRs develop in the first wave, followed by amacrine cells, bipolar cells, MG and rod PRs in the second.[5] These developmental waves have also been observed in retinal

organoids (rORGs) (Fig 1b).[6,7] In RD diseases such as AMD, PRs can die from a lack of nutrients due to loss of support from the RPE,[8] whereas in glaucoma, RGCs experience axonal damage and apoptosis, thus disrupting communication to the brain.[9] With the understanding that these cells play critical roles in RD progression, creating systems to understand the developing human retina *in vitro* is a pressing need.

One promising goal for RD research is to harness the potential of regeneration to combat cell death in RD diseases. MG are non-neuronal cells that span the width of the retina and are critical in maintaining homeostasis of the eye. In many non-mammalian species, MG respond to retinal injury by undergoing cell proliferation and gaining stem cell-like identity in order to replenish the depleted neurons.[4] Although MG do not exhibit these proliferative capabilities in humans, human induced pluripotent stem cells (iPSCs) can be used to investigate the transcriptomic patterns necessary for the differentiation of specific cell types. iPSCs are stem cells generated through the reprogramming of somatic cells to revert them back to a pluripotent state.[10] Understanding the transcriptomic patterns necessary for iPSCs to differentiate reliably to specific cell types could pave the way for inducing MG-directed regeneration of cells in the human retina. This would be desirable as a much less invasive clinical technique compared to stem cell injections and prosthetics.

While animal models have been helpful for studying many retinal diseases,[11–14] the complex genetic nature of RD renders these models suboptimal for elucidation of underlying mechanisms of human development and disease.[15] As such, *in vitro* models such as iPSC-derived organoids have gained popularity because of their human background. Concurrently, rapid progress in the field of bioinformatics has allowed for

many new insights into the developing human retina. Single-cell RNA sequencing (scRNA-seq), in particular, has led to new high-resolution insight into the way distinct cell populations within the retina change over time.[16–19] Even more recently, advances in machine learning (ML) and artificial intelligence (AI) have begun to pave the way for new and efficient ways of processing data.[20,21] In Python, the scvi-tools package can be used for many types of probabilistic analyses using neural networks,[22] such as integration of data[23] and label transfer.[24] These kinds of new tools allow researchers to derive more information from biological data, alleviate technical noise and uncover true biological trends hidden beneath the surface.

In this thesis, I examine PRs purified from iPSC-derived rORGs (O-PRs). rORGs are formed by aggregating iPSC cells and allowing them to spontaneously differentiate and develop into the maturing retina, with notable features like the PR outer segment and RPE. The presence of a *CRX*-mRuby3 reporter built into the original iPSCs allows PRs to be identified based on fluorescence (Fig 1c). For this project, the purification of O-PRs was performed by fluorescent activated cell sorting (FACS) with mRuby3 red fluorescence (Fig 1d). Purified O-PRs were processed for scRNA-seq (Fig 1e) at days 90 and 120 to validate the cellular identity of these cells. These timepoints were chosen because they represent a critical juncture in rod and cone PR development, wherein cones have begun maturing and rods have just started to appear. Analyzing multiple timepoints allows for a more in-depth analysis where development of O-PRs can be explored through elucidation of trends in gene expression patterns and functional pathways. Additionally, understanding the patterns and key developmental timepoints of PRs, including rod and cones, in rORGs could lead to

a better understanding of possible MG trans-differentiation strategies, contributing to the long-term goal of rescuing RD disease phenotypes in humans.

Analysis of O-PRs by scRNA-seq (Fig 1f) confirmed their PR-like identity. I found that cone marker genes (e.g. *PDE6C, ARR3*) were expressed convincingly in day 90 and 120 O-PRs, while rod marker genes (e.g. *NR2E3, NRL*) did not show significant expression until day 120. These observations were in line with the timeline of rod and cone PR development (Fig 1b). Importantly, once expressed, both rods and cones were present in distinct clusters. O-PRs were also compared to published *in vitro* and *in vivo* datasets through integration and ML-based label transfer, which further confirmed PR-like identity and developmental trajectory.

Along with the explosion in popularity with single-cell technology comes a need for reproducibility in publications utilizing the various tools being rapidly developed. Recently, the National Institutes of Health (NIH) introduced the Rigor and Reproducibility policy for grant submissions,[25] and high-impact journals such as Cell and Nature have adopted the FAIR (Findable, Accessible, Interoperable, Repeatable) standards[26] for the articles they publish. A secondary goal of this work is to provide a template for the computational analysis of human cell-type models, retinal or otherwise. Towards the aforementioned goals of reproducibility, all the code used in the analyses presented in this thesis has been documented and published on GitHub with the goal of enabling scientists without a bioinformatics background to perform their own analysis on their data[27].

Materials & Methods

Retinal Organoid Culture

Organoids were grown using a modified version of previously published protocols.[28,29] Briefly, iPSCs with the endogenous *CRX*-mRuby3 and *ARR3*-TagBFP2 reporters were grown to 70% confluence and dissociated to single cells, counted, and seeded into each well of a non-TC 96-well U-bottom plate at 2,000 cells per well in mTeSR plus 2µM Thiazovivin. Aggregates formed overnight and were then gradually transitioned into 100% Neural Induction Media (NIM) from day 1 to day 3. NIM consisted of DMEM/F12 (Life technologies catalog #11330-057), 1% N2 supplement, 1% MEM non-essential amino acids, 1% anti-anti, and heparin (2 µg/ml). On day 6, the NIM was supplemented with 1.5nM BMP-4. On day 8, organoids were plated on FBS-coated plates and allowed to attach and develop until day 16, while BMP4 was gradually depleted. On day 16, organoids are lifted from the plate, and transferred into a 1% FBS Retinal Differentiation Media (RDM). RDM consisted of DMEM/F12 (3:1), 2% B27 supplement, 1% MEM non-essential amino acids, and 1x anti-anti) with 1% fetal bovine serum (FBS). From days 16 to 25, the FBS concentration was gradually increased to 10%. From day 25, organoids were fed with Long Term Retinal media (LTR). LTR consisted of a 3:1 mix of DMEM (Invitrogen #11965): F12 (#11765: Invitrogen) with 1% B27 (#17504044: Invitrogen), 10% heat inactivated FBS (#16140071; Invitrogen), 1 mM pyruvate (#11360; Invitrogen), 1xNEAA (#11140, Invitrogen), 1X Glutamax (#35050061; Invitrogen) and 1 mM taurine (#T-8691; Sigma). From days 50 to 120, the LTR was supplemented with All-Trans Retinoic Acid (ATRA).

Fluorescence Activated Cell Sorting (FACS) of Retinal Organoids

iPSC-derived rORGs with the built in endogenous *CRX*-mRuby3 and *ARR3*-TagBFP2 reporters were grown to timepoints day 90 and 120 for FACS. At these timepoints, organoids were dissociated using papain and loaded into a Biorad S3 FACS sorter to isolate CRX+ PRs and CRX+/ARR3+ cones. The *ARR3*-TagBFP2 reporter did not produce a pure population of cone PRs, likely due to the immature nature of rod and cone PRs at days 90 and 120. For further analysis, all the FACS samples were pooled, yielding a *CRX*-mRuby3 positive PR population with both rods and cones.

Single-cell RNA Sequencing (scRNA-seq) of O-PRs

scRNA-seq of *CRX*+ FACS-purified O-PRs was performed at organoid days 90 and 120 (n=1, 2) to investigate the crucial timepoint when PRs emerge and begin to mature. For each replicate, 5,000 cells were collected in 5uL of cell resuspension buffer for particle-templated instant partition sequencing (PIP-seq)[30]. Cells were captured, lysed and libraries were prepared according to manufacturer's instructions (PIPseq™ T2 3' Single Cell RNA Kit v4.0). Briefly, mRNA was captured from samples via addition of breaking buffer to disrupt the stable emulsion, and cDNA was amplified and isolated from the PIPs by SPRI magnetic bead purification. The quality and quantity of cDNA was assessed using a 4150 TapeStation with the D5000 ScreenTape (Agilent Technologies). Samples with >10ng cDNA and peaks within the acceptable range were carried forward for library preparation. Fragmentation, end repair, A-tailing and adapter ligation were carried out

followed by sample index PCR with unique dual index combinations. Sequence-ready libraries were again tested for quality and quantity on the TapeStation with a high-sensitivity D1000 ScreenTape. Sequencing was performed on a NovaSeq 6000 (100-bp paired-end reads, Illumina) at the UCSD Institute for Genomic Medicine (IGM) Core to obtain 150-200 million reads per sample.

scRNA-seq Analysis of O-PRs

Processing of raw reads was carried out using Pipseeker (v3.0.5, FluentBiosciences) to obtain filtered count matrices. Briefly, barcodes and molecular identifiers were assigned to reads, reads were mapped to the reference genome using STAR,[31] transcripts were quantified, and cell calling was performed using a knee plot at multiple sensitivity levels. For each sample, cell calling was checked manually to ensure that the tradeoff between number and quality of cells was acceptable. After Pipseeker processing, filtered count matrices were loaded into R (Seurat v4.4)[32] or Python (Scanpy v1.10.1)[33], and basic filtering and preprocessing was carried out. Briefly, barcodes were made unique, outliers in total RNA and cells with high mitochondrial gene expression (> 5%) were filtered out. Next, dimensionality reduction was performed by first calculating principal components and then using uniform manifold approximation and projection (UMAP) to visualize the cells in 2D space. Additionally, clustering was performed with the Louvain algorithm to color cells that were most similar to each other. From this point, various analyses were carried out, including gene expression analysis, pathway analysis, integration and label transfer.

Gene Expression and Pathway Analysis

Gene expression and pathway analysis was done with Seurat (v4.4). Gene expression was visualized with the DotPlot function. Additionally, UMAPs were plotted using the DimPlot and FeaturePlot functions to highlight Louvain clusters and expression of specific genes in each cell, respectively. Retinal progenitor cells (RPCs) were identified in two Louvain clusters based on their lack of expression of typical PR markers (e.g. *CRX*, *NEUROD4*) and expression of immature RPC markers (e.g. *HES5*, *PAX6*). Pathway analysis was then performed between different groups (retinal progenitors, day 90 O-PRs and day 120 O-PRs) by first generating lists of group-specific markers with the FindMarkers function. The resulting markers for each group were sorted by their log2FC to create a ranked list of genes that were differentially expressed in that group of interest. From there, gene set enrichment was performed with the gseGO function from the clusterProfiler (v4.12.0)[34] package to obtain top enriched pathways for each group, which were then plotted with barplots sorted by net enrichment score (NES). Barplots were generated in Python (v3.9) using the seaborn (v0.13.2) barplot function.[35] Heatmap of select genes from each group was plotted by generating pseudobulk data with the AverageExpression function.

Integration Analysis

Integration was carried out using Seurat (v4.4). First, the SelectIntegrationFeatures function was used to select features that showed high variance in each dataset. Next,

anchors for each dataset are found using FindIntegrationAnchors. These anchors represent pairs of cells across the two datasets which are determined to represent a similar biological state. Selecting them as anchors allows for harmonization of the data, which is done using the final function IntegrateData, which performs integration based on the anchor set. After these steps, the integrated data is processed as normal by performing scaling, dimensionality reduction and clustering, and then visualized on a UMAP colored for the original sample to determine overlap. This overlap can show us which cells in each dataset are similar to each other, making it a useful tool for comparisons between datasets.

Label Transfer Analysis of day 120 O-PRs

Label transfer was done in Python (v3.9) with scanpy using the scvi-tools package (v1.1.2)[22]. Briefly, day 120 O-PRs and previously published reference datasets were loaded, and reference data was subsetted to keep only rod and cone PRs. The data was then filtered for outliers, including total genes expressed and mitochondrial gene percentage, and the reference data was concatenated with day 120 O-PRs. The resulting anndata object was setup for label transfer using the model.SCVI.setup_anndata function. Next, a variational autoencoder (VAE) model was generated from the anndata object using the model.SCVI function, and this model was trained using the train function. The function model.SCANVI.from_scvi_model was then used to generate a VAE from the datasets, and finally after another training round labels were predicted for day 120 O-PRs using the predict function. UMAPs were used to visualize the unlabeled versus labeled day 120 O-PRs, as well as the original references datasets

Data and Code Availability

Single-cell data matrices from day 90 and 120 O-PRs as well as all code used to generate figures is publicly available in a GitHub repository.[36] All external datasets used in this work can be found in the references of this manuscript as well as in the same GitHub repository.

Results

O-PRs Display PR-like Identity

To assess the transcriptional identity of organoid-derived PRs (O-PRs), scRNA-seq was carried out on O-PRs sorted from both day 90 and 120 rORGs using *CRX* as a reporter for PRs. Cells were clustered using the Louvain algorithm and projected into two-dimensional space by a uniform manifold approximation and projection (UMAP) to visualize the clustering pattern and gene expression of O-PRs (Fig 2a,b). Cell type marker analysis across all clusters revealed that most clusters displayed strong neuronal (*DLG4*, *MAP2*, TUBB3)[37] and PR (*CRX*, *OTX2, NEUROD4, RCVRN*)[37–39] marker gene expression. Markers for glial (*S100B, PDGFRA, OLIG2*)[37], amacrine (*GAD1, TFAP2B, TFAP2C*)[37], horizontal (*LHX1, CALB1, PTF1A*)[37], bipolar (*NETO1, PCP2*)[37,40] and RGC (*POU4F1, SNCG, NEFL*)[37] were not expressed to any meaningful levels, as expected. This confirmed the purity of the O-PR population. Notably, clusters 8 and 9, made up of cells from both day 90 and 120 O-PRs, expressed retinal progenitor cell (RPC) markers *VIM, PAX6, SOX2* and *VSX2* (Fig 2b).[41,42] *VSX2* was also observed to a lesser degree in clusters 1, 5 and 6, indicating that some cells in these clusters were at some state between an RPC and a committed PR. Additionally, cluster 10 clustered closer to the RPC than the PR clusters on the dotplot. Cluster 10 did express small amounts of PR and RPC markers, but its identity was unclear because of low marker expression across the board, as well as a low number of cells (22 cells, or only 0.58% of the total). It is possible that it represents a different group of RPCs than those present in clusters 8 and 9.

To further explore this progenitor identity, gene set enrichment analysis (GSEA) was carried out to identify biological pathways that were upregulated in clusters 8 and 9. Genes were ranked by their differential expression in clusters 8 and 9 versus all other clusters, and the top 10 pathways by net enrichment score (NES) were visualized with a bar graph. Top pathways included many proliferation and developmental processes, such as regulation of neural precursor cell proliferation (GO:2000177), regulation of cell population proliferation (GO:0042127) and tube morphogenesis (GO:0035239) (Fig 2c). Similar analysis was conducted on day 90 O-PRs versus RPCs and day 120 O-PRs vs RPCs to assess enriched biological mechanisms in the two timepoints. Top pathways enriched in day 90 O-PRs included many neuronal mechanisms related to synaptic activity and chemical transmission, while the top day 120 O-PR mechanisms showed terms more indicative of photoreceptors, including sensory perception of light stimulus (GO:0050953) and visual perception (GO:0007601), showing a clear increase in PR-like activity from day 90 to 120 (Fig 2c). Looking past the top 10 pathways in each group showed that while not as significantly enriched as in day 120, day 90 O-PRs still showed upregulation of photoreceptor-specific pathways. Enrichment plots for the detection of light stimulus (GO:0009583) term were generated for RPCs, day 90 O-PRs and day 120 O-PRs. While the pathway was downregulated in RPCs, both day 90 and 120 O-PRs showed significant enrichment of the process, with day 120 showing a larger presence of genes in the gene set, especially towards the top of the ranking. Ten functionally important genes for PR development (e.g. *PRPH2, RCVRN*, *ROM1*) from the detection of light stimulus pathway were plotted in a heatmap to highlight the upwards trend of genes involved in visual

perception by averaging gene expression of cells in each group (Fig 2d). This scRNA-seq data thus showed convincing evidence that day 90 and 120 O-PRs display a strong PR-like identity, using marker gene expression and pathway analysis.

Many types of RD specifically affect one PR subtype more severely. For example, RP often leads to night blindness and loss of peripheral vision, symptoms associated with rod PR degeneration.[43] Given the importance of rod and cone PR subtypes, I next set out to investigate the ability of O-PRs to recapitulate this subtype diversity found in typical PRs.

O-PRs Show Signs of Rod and Cone Photoreceptor Subtypes

UMAPs were once again generated and visualized for cell populations (day 90/120 RPCs, day 90 O-PRs and day 120 O-PRs) as well as Louvain clusters (Fig 3a). Gene expression of cone (*GUCA1A, PDE6C, GNAT2, ARR3*) and rod (*NR2E3, NRL, PDE6B, GNAT1*) markers were plotted in a dotplot across all Louvain clusters. Clusters 0, 2 and 4 showed strong expression of cone markers, while clusters 6 and 7 showed the strongest rod marker expression. Out of these 5 clusters, only cluster 0 was from the day 90 O-PRs, indicating that most clusters expressing specific cone or rod markers were from the more mature day 120 O-PR group (Fig 3b). Additionally, almost all cells expressing significant levels of rod markers were from the day 120 group, which aligns with the developmental timeline of the retina, in which cones develop first and rods begin to develop only around day 100. To elucidate the maturity of the rod and cone populations in O-PRs, a UMAP was plotted, and cells were colored according to how many of the 4 cone or rod marker genes they expressed. Arrows were drawn on the UMAP to highlight maturation trends. These

trends indicate that the rods and cones present in O-PRs have not arrived at a steady state and are likely to still be maturing (Fig 3c,d). In a fully mature retina, the gene expression would reach this 'steady state' where distinct trends would not be observed. Notably, there was significantly less expression of rod markers overall compared to the cone markers, which is consistent with the developmental timeline, in which cones begin developing earlier than rods. More mature timepoints up to and past day 200 would likely be necessary for full maturation of rod and cone PRs. This is consistent with previous studies that show the transcriptome of rORGs do not stabilize until as late as days 210-270.[18]

Up until this point, I have assessed the PR-like identity of O-PRs through investigations centered around expression of specific cell type markers. While this is a good benchmark, comparison of O-PRs to other published datasets, especially those generated from *in vivo* methods (fetal and adult donors) offer the best tools for assessing validity of induced cell types. In fact, the fetal retina has been found to closely match the developmental steps of rORGs.[39]

O-PRs Integrate Well with *in vivo* Datasets

In order to test the compatibility of O-PRs with human fetal data, recently published scRNA-seq gene expression matrices were obtained at fetal timepoints day 82 and 105,[39] to be compared with day 90 and 120 O-PRs, respectively. Day 90 O-PRs integrated well with the PR population from day 82 fetal retina, and day 120 O-PRs integrated similarly with day 105 fetal retina, with slightly less overlap possibly due to the larger difference in the timepoint (8 days versus 15 days) (Fig 4a). PR populations were

determined by expression of markers *OTX2* and *RCVRN*, as was done in the original study.[39] Overlap and nearby clustering of cells from the two datasets tells us about the similarity of the two groups. Dataset integration occurs through the elucidation of integration anchors, pairs of cells that represent similar biological states across the groups. Using these anchors, I harmonized the datasets together and plotted them. After these steps, similar cells will cluster very close to each other, or even overlap. Thus, this integration analysis was performed to show that O-PRs at both day 90 and 120 are similar to fetal PRs at similar timepoints. However, the integration does not tell us anything regarding the cone and rod PR subpopulations present in O-PRs since the reference dataset was not labeled for rod and cone subtypes.

Reference Datasets Highlight both Rod and Cone PR Populations in day 120 O-PRs

Rod and cone O-PR populations were investigated through marker gene expression and resulted in the elucidation of distinct clusters. However, this analysis is limited by the small number of genes analyzed. Label transfer analysis takes a pre-labeled reference dataset and transfers these labels onto unlabeled experimental data using a ML model. Since this ML model is trained on all dimensions of the reference data, using this strategy to identify rod and cone populations in our dataset compliments marker gene analysis by utilizing cells' entire gene expression vectors. In order to confirm that day 120 O-PRs contain distinct rod and cone PR populations, reference data from iPSC-derived rORG PRs[18] and adult human PRs[44] were both used to separately label day 120 O-PRs using a machine learning model. Label transfer with both references resulted in similar

classification of rod and cone PRs in day 120 O-PRs (Fig 4b,c), showing that both cell types in day 120 O-PRs mimic *in vitro* and *in vivo* datasets further confirming their validity as a model for human PRs. Additionally, this result implied that rods and cones contain similar characteristics in *in vitro* and *in vivo* models, although a more comprehensive study focused on this would need to be carried out to confirm.

Discussion

O-PRs as a Model for Human PRs

In this work I analyzed O-PRs to assess their suitability as *in vitro* models of human PRs. Using multiple analytical methods, I concluded that O-PRs do recapitulate PR identity. This was shown through analysis of the transcriptomic patterns of the O-PRs, as well as investigation of enriched biological pathways based on differentially expressed genes. Through similar analysis, I also found that the O-PRs showed distinct rod and clone clusters. Rod specific marker genes such as *NRL* and *NR2E3* have been shown to turn on around the day 100 mark in organoids[6] and around day 80 in fetal retina,[45,46] which is consistent with the fact that there was not a significant level of rod marker (e.g. *NRL*, *NR2E3*) expression apparent in day 90 O-PRs, while day 120 O-PRs showed some clusters with significant expression of these markers. Cones, on the other hand, are known to develop earlier[5] and as such showed strong marker (e.g. *GUCA1A*, *ARR3*) expression in both day 90 and 120 O-PRs. In general, cells with cone markers were more abundant than those with rod markers, even when looking only at day 120 O-PRs.

Day 90 and 120 O-PRs reflect human PRs at comparable stages in development, but importantly, they do not recapitulate fully mature PRs, nor did I expect them to. This lack of maturity was evident when looking at the upwards trend in PR-specific functional pathways, as well as the progression of specific pathways through the RPC, day 90 O-PR and day 120 O-PR groups. A lack of maturity was additionally evident in pseudotime analysis of rod and cone markers. Pseudotime analysis is a popular method for inferring a developmental trajectory of cells at some singular timepoint,[47] and it works by identifying

key ‘ordering genes’ and generating a trajectory based on trends in their expression. In this work, I performed a much-simplified version of this kind of analysis by visualizing co-expression of four rod and four cone markers using UMAPs. If O-PRs were fully mature, I would expect to see a ‘steady-state’, wherein no trends would be observed within clusters. However, the plots indicate clear trends of less-to-more marker gene expression, confirming the ongoing maturation at this timepoint.

Another indication that day 90 and 120 O-PRs have not finished development was the presence of a distinct RPC cluster in the data. Retinal cell differentiation from RPCs is a complex topic. Multiple methods of differentiation have been proposed, mainly: multi-potent RPCs whose competency changes intrinsically with time to allow for generation of different cell types, fate-restricted RPCs with intrinsic patterning that give arise to specific retinal subtypes and multipotent RPCs whose fates are determined by extrinsic environmental cues.[48] Growing evidence, however, lends to the conclusion that the true mechanism is that RPCs go through different states of competency which allow to give rise to specific cell types.[49–52] In agreement with this model, we see RPC clusters in both day 90 and 120 O-PRs, indicating that through these two timepoints a small population of RPCs continues to be present, further confirming the conclusion that day 120 O-PRs have not reached maturity. Interestingly, expression of some of the RPC markers used (e.g. *VIM, VSX2*) was detected to some degree in clusters outside of 8 and 9, possible highlighting RPCs at various stages of differentiation. Why RPCs were detected in our dataset is not completely clear since cells were based on FACS sorted CRX-mRuby3+ expression. One possibility is that CRX negative cells were not gated as stringently as possible. On the

other hand, some of the progenitors did express *CRX* which calls into question how mature CRX+ cells are at the time of collection. Regardless, the presence of these RPCs is useful for the interpretation of the state of day 90 and 120 O-PRs and strengthens the conclusion that O-PRs are indeed a good model for PRs.

In summary, in this thesis I show that O-PRs do capture PR-identity, including the presence of rods and cones, and that at day 90 and 120 they are still developing and not fully mature. This also implies that rORGs at this timepoint likely lack many other late-born retinal cells such as MG and bipolar cells.

Purification of O-PRs

While previous work has been carried out analyzing PRs in whole retinal organoids,[18,39] my work importantly analyzes O-PRs that have been physically purified from non-PR retinal cells before sequencing and subsequent analysis. The ability to purify these PRs with the *CRX*-mRuby3 cassette is an important step towards studying PRs *in vitro*. Although purification of the PR population was successful, the purification of rod and cone PRs as separate fractions using an endogenous *ARR3*-tagBFP cassette did not yield completely pure populations. Though *ARR3* was expressed in cone clusters, some mixed expression due to the immaturity of the PR subtypes likely contributed to its non-specificity. It is possible that upon repetition of this experiment with later, more mature timepoints, the *ARR3*-TagBFP2 reporter would allow for FACS separation of rods and cones.

Future Induced PR Studies and Retinal Organoids as a Model for Regeneration

The long-term goal for rORGs is to use them as a drop-in replacement for the *in vivo* human retina. *In vitro* trans-differentiation of MG into specific cell types that have been lost to injury is the holy grail of retinal regeneration and is a highly complex task. The entire process can be divided into two main steps: dedifferentiation of MG to a proliferative, stem-cell like state and the differentiation of these cells into the cell type of interest, in this case, PRs. Many groups have been working on the first step by studying the proliferative potential of MG in different models.[53–57] On the other hand, this work presents insight on the second step of the process, wherein the proliferative cells must differentiate into PRs. Understanding the developmental trajectories and transcriptomic patterns of O-PRs could conceivably aid in the elucidation of differentiation pathways. However, for these O-PRs to be fully established as a good model more timepoints beyond day 90 and 120 must be studied to get a complete picture of the maturation of these O-PRs into late-stage maturity. Additionally, further analysis can be done on the data generated from the day 90 and 120 O-PRs regarding functional activity of the PRs, and *de novo* discovery of important genes. Once these necessary steps have been taken, experiments on overexpression of key genes to push proliferating MG into a PR-fate could begin to take place.

scRNA-seq as a Viable Technology

As with many studies in the field of retinal development, and especially retinal organoid-based studies, this work relies entirely on scRNA-seq for its analysis. While it is

a powerful tool with many capabilities, like many technologies, one leg can be hard to stand on. Compared to traditional (bulk) RNA-seq, scRNA-seq poses a problem of significantly lower sequencing depth. Whereas a typical RNA-seq dataset could garner 20-50+ million reads per sample, single cell sequencing studies often fall around 100,000 reads per cell. This significant drop in coverage means that relatively lowly expressed genes that would make it into an RNA-seq dataset could get dropped from a scRNA-seq dataset, perhaps due to bias introduced during amplification.

Additionally, problems may arise for datasets with a homogenous population. With a highly similar dataset, like a pure population of one cell type, the similarities between cells often outweigh the differences, leading to unexpected clustering behavior unhelpful to answering questions about the biological state of the cells. While this study did contain a mostly pure population of PRs, the rod and cone subtypes were distinct enough to add some heterogeneity to the dataset. In a recently published study on induced RGCs (RGC-iNs),[37] scRNA-seq was used to successfully look at the developmental trajectory of RGC-iNs through three weeks. The scRNA-seq was also able to unveil some details about the types of subtypes that may be present. However, unlike PRs, RGC subtypes are not as well understood, and have many overlapping markers and transcriptomic profiles. As such, attempts to extract distinct cell type populations in RGC-iNs was inconsistent. On the other hand, highly heterogenous datasets like whole-organoid data produce extremely relevant clusters because of the relatively large gap in transcriptomic profiles. Researchers should be cautious when planning experiments and deciding if scRNA-seq alone is the best approach. Often, the best option is using both RNA-seq and scRNA-seq to compliment and

enrich each other. On this note, a key step in bolstering the work done in this study will be to perform bulk-RNA-seq to add confidence to the conclusions put forth here.

Figures

Figure 1: Experimental overview

(A) Drawing depicting the layered anatomy of the retina. **(B)** Overview of the *CRX*-mRuby3 reporter system and its PR-specific labeling. **(C)** Developmental timeline of retinal organoids, including four main phases. Timepoints sequenced in this study are shown in red underneath the timeline. **(D)** Overview of fluorescence activated cell sorting (FACS) workflow used to sort out *CRX*-mRuby3 positive cells, resulting in a pure PR population. **(E)** Particle templated instant partition sequencing (PIPSeq) workflow by which sorted PRs were prepared for sequencing. **(F)** Schematic depicting the single-cell RNA sequencing workflow.

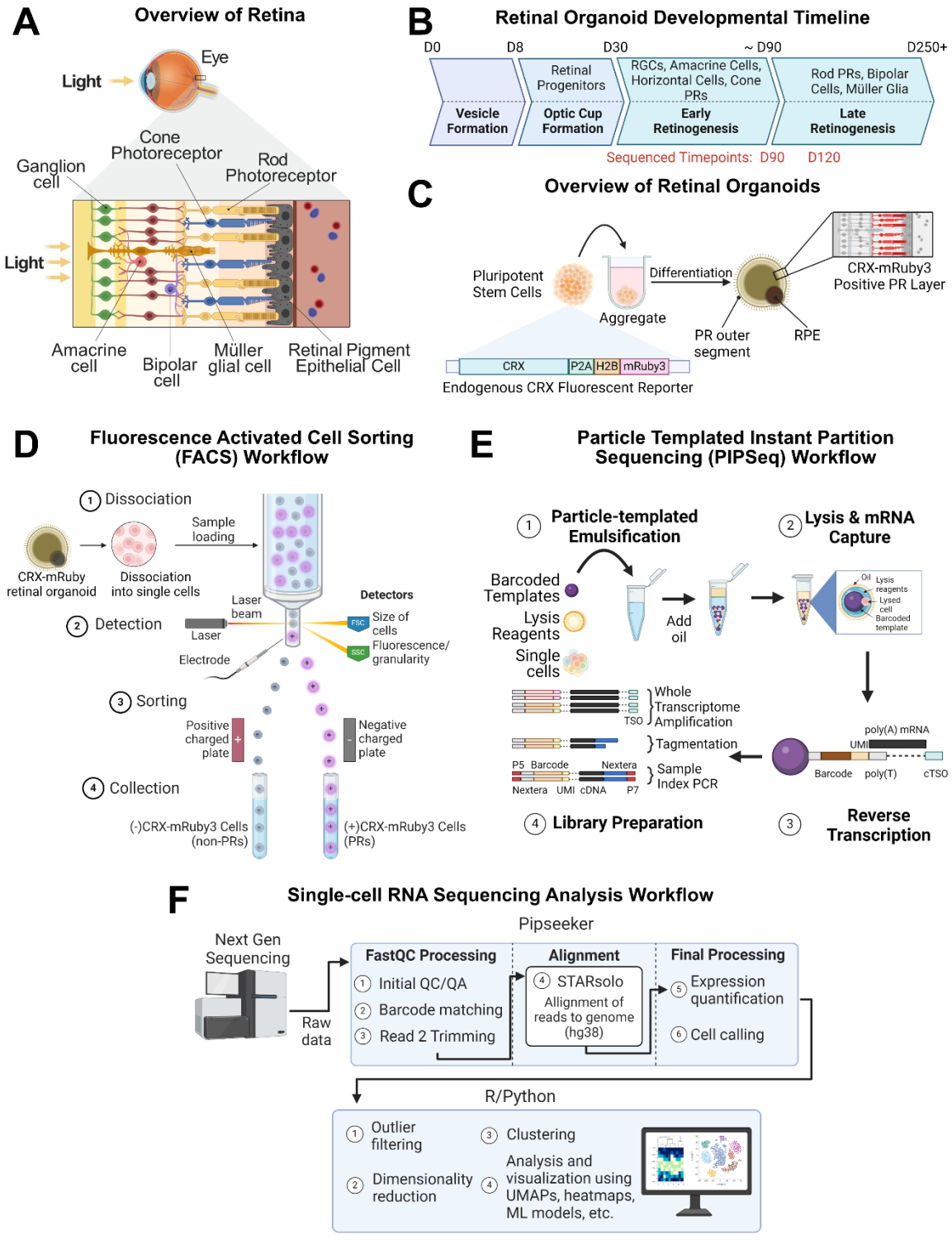
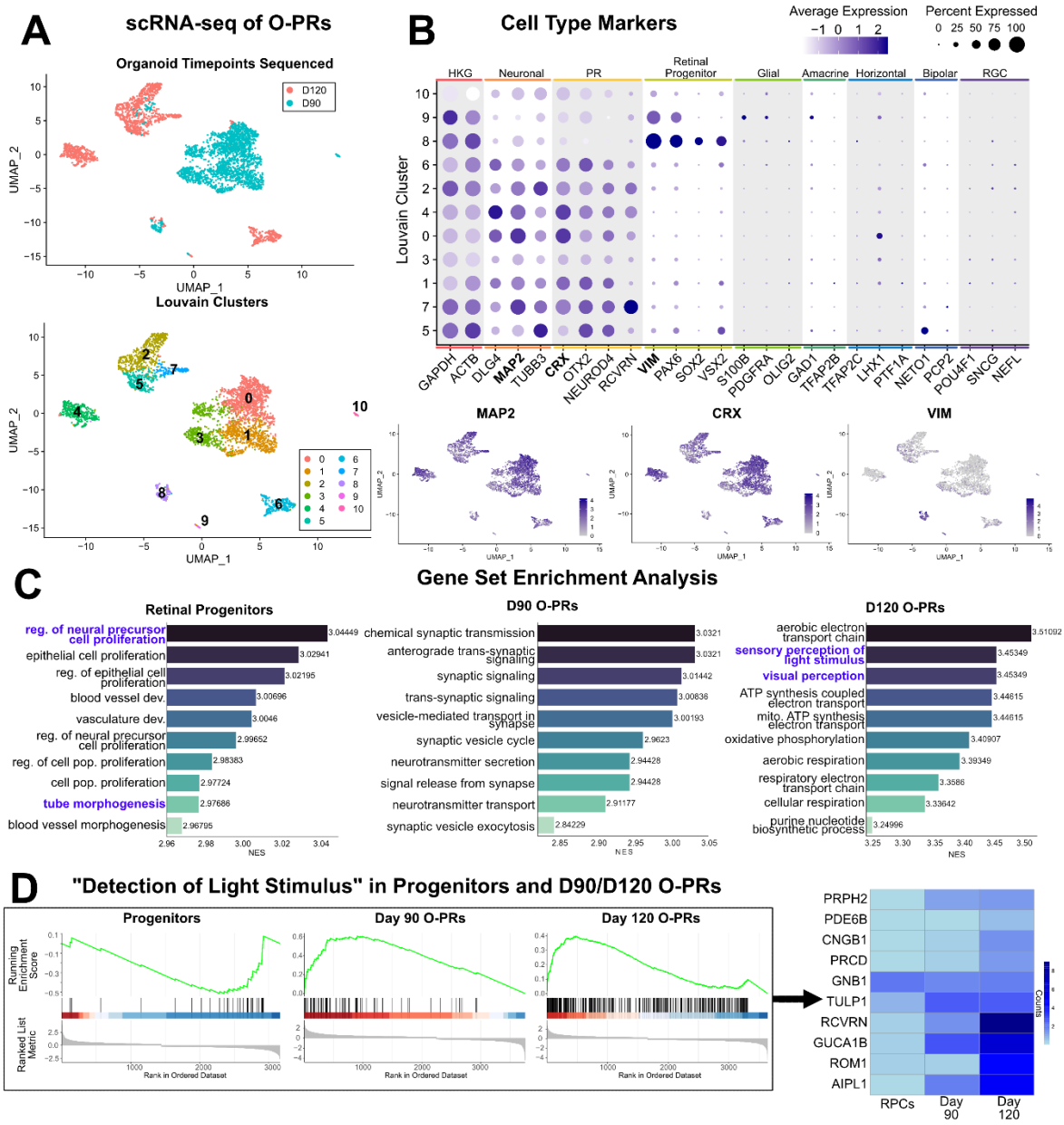

A
Overview of Retina
Light
Eye
Cone Photoreceptor
Rod Photoreceptor
Ganglion cell
Light
Amacrine cell
Bipolar cell
Müller glial cell
Retinal Pigment Epithelial Cell
B
Retinal Organoid Developmental Timeline
D0
D8
D30
~ D90
D250+
Vesicle Formation
Optic Cup Formation
Retinal Progenitors
Early Retinogenesis
RGCs, Amacrine Cells, Horizontal Cells, Cone PRs
Late Retinogenesis
Rod PRs, Bipolar Cells, Müller Glia
Sequenced Timepoints: D90 D120
C
Overview of Retinal Organoids
Pluripotent Stem Cells
Aggregate
Differentiation
CRX-mRuby3 Positive PR Layer
PR outer segment
RPE
CRX
P2A
H2B
mRuby3
Endogenous CRX Fluorescent Reporter
D
Fluorescence Activated Cell Sorting (FACS) Workflow
1 Dissociation
CRX-mRuby retinal organoid
Dissociation into single cells
Sample loading
2 Detection
Laser beam
Laser
Electrode
Detectors
Size of cells
Fluorescence/ granularity
3 Sorting
Positive charged plate
Negative charged plate
4 Collection
(-)CRX-mRuby3 Cells (non-PRs)
(+)CRX-mRuby3 Cells (PRs)
E
Particle Templated Instant Partition Sequencing (PIPSeq) Workflow
1 Particle-templated Emulsification
Barcoded Templates
Lysis Reagents
Single cells
Add oil
2 Lysis & mRNA Capture
Oil
Lysis reagents
Lysed cell
Barcoded template
3 Reverse Transcription
poly(A) mRNA
UMI
Barcode
poly(T)
cTSO
4 Library Preparation
TSO
Whole Transcriptome Amplification
Tagmentation
P5
Barcode
Nextera
Sample Index PCR
Nextera
UMI
cDNA
P7
F
Single-cell RNA Sequencing Analysis Workflow
Next Gen Sequencing
Raw data
Pipseeker
FastQC Processing
1 Initial QC/QA
2 Barcode matching
3 Read 2 Trimming
Alignment
4 STARsolo
Allignment of reads to genome (hg38)
Final Processing
5 Expression quantification
6 Cell calling
R/Python
1 Outlier filtering
2 Dimensionality reduction
3 Clustering
4 Analysis and visualization using UMAPs, heatmaps, ML models, etc.

Figure 2: Single cell sequencing of O-PRs reveals PR-like identity.

(A) UMAPs of O-PRs colored by timepoint (top) and louvain clustering (bottom). **(B)** Dotplot indicating expression of select cell type marker genes in each louvain cluster (top) and UMAPs displaying expression of neuronal (*MAP2*), photoreceptor (*CRX*) and retinal progenitor (*VIM*) markers. **(C)** Barplots depicting top 10 gene ontology pathways by net enrichment score (NES), generated from list of genes ranked by their differential expression in the putative progenitor clusters (8, 9) versus all other clusters (left), day 90 O-PRs versus all other clusters (middle) and day 120 O-PRs versus all other clusters (right). **(D)** GSEA enrichment plots of the Detection of Light Stimulus pathway (GO:0009583) across progenitors, day 90 O-PRs and day 120 O-PRs (left) and heatmap of normalized counts of select genes from the pathway


A
scRNA-seq of O-PRs
Organoid Timepoints Sequenced
D120
D90
UMAP_2
UMAP_1
Louvain Clusters
B
Cell Type Markers
Average Expression
Percent Expressed
HKG
Neuronal
PR
Retinal Progenitor
Glial
Amacrine
Horizontal
Bipolar
RGC
Louvain Cluster
GAPDH
ACTB
DLG4
MAP2
TUBB3
CRX
OTX2
NEUROD4
RCVRN
VIM
PAX6
SOX2
VSX2
S100B
PDGFRA
OLIG2
GAD1
TFAP2B
TFAP2C
LHX1
PTF1A
NETO1
PCP2
POU4F1
SNCG
NEFL
MAP2
CRX
VIM
C
Gene Set Enrichment Analysis
Retinal Progenitors
reg. of neural precursor cell proliferation
epithelial cell proliferation
reg. of epithelial cell proliferation
blood vessel dev.
vasculature dev.
reg. of neural precursor cell proliferation
reg. of cell pop. proliferation
cell pop. proliferation
tube morphogenesis
blood vessel morphogenesis
NES
D90 O-PRs
chemical synaptic transmission
anterograde trans-synaptic signaling
synaptic signaling
trans-synaptic signaling
vesicle-mediated transport in synapse
synaptic vesicle cycle
neurotransmitter secretion
signal release from synapse
neurotransmitter transport
synaptic vesicle exocytosis
NES
D120 O-PRs
aerobic electron transport chain
sensory perception of light stimulus
visual perception
ATP synthesis coupled electron transport
mito. ATP synthesis electron transport
oxidative phosphorylation
aerobic respiration
respiratory electron transport chain
cellular respiration
purine nucleotide biosynthetic process
NES
D
"Detection of Light Stimulus" in Progenitors and D90/D120 O-PRs
Progenitors
Day 90 O-PRs
Day 120 O-PRs
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
PRPH2
PDE6B
CNGB1
PRCD
GNB1
TULP1
RCVRN
GUCA1B
ROM1
AIPL1
Counts
RPCs
Day 90
Day 120

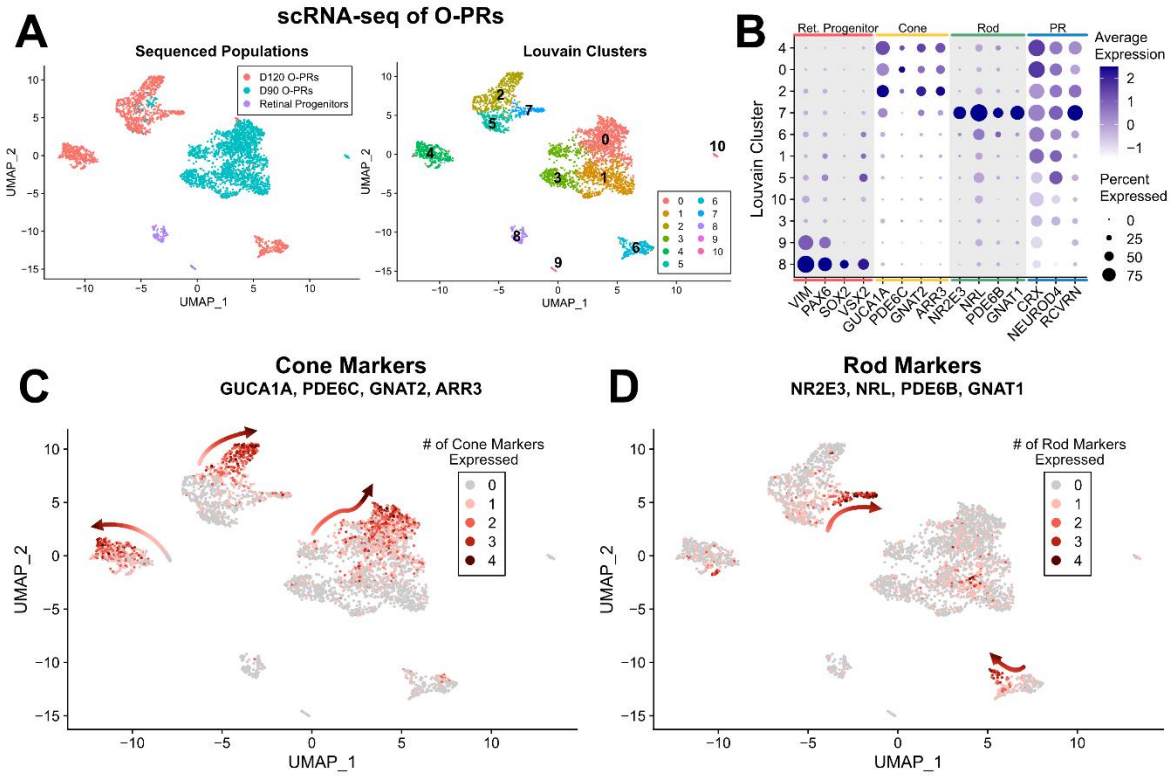


Figure 3: Day 90 and 120 O-PRs show evidence of maturing rod and cone populations.

(A) UMAPs of O-PRs colored by population (left) and Louvain clustering (right). **(B)** Dotplot indicating expression of retinal progenitor, cone, rod and photoreceptor markers in each louvain cluster. **(C)** UMAP highlighting the overlap of select cone markers (*GUCA1A*, *PDE6C*, *GNAT2*, *ARR3*). **(D)** UMAP highlighting the overlap of select rod markers (*NR2E3*, *NRL*, *PDE6B*, *GNAT1*). **(C-D)** Genes were considered expressed if they had normalized count > 1. Arrows indicate maturation trends within clusters.

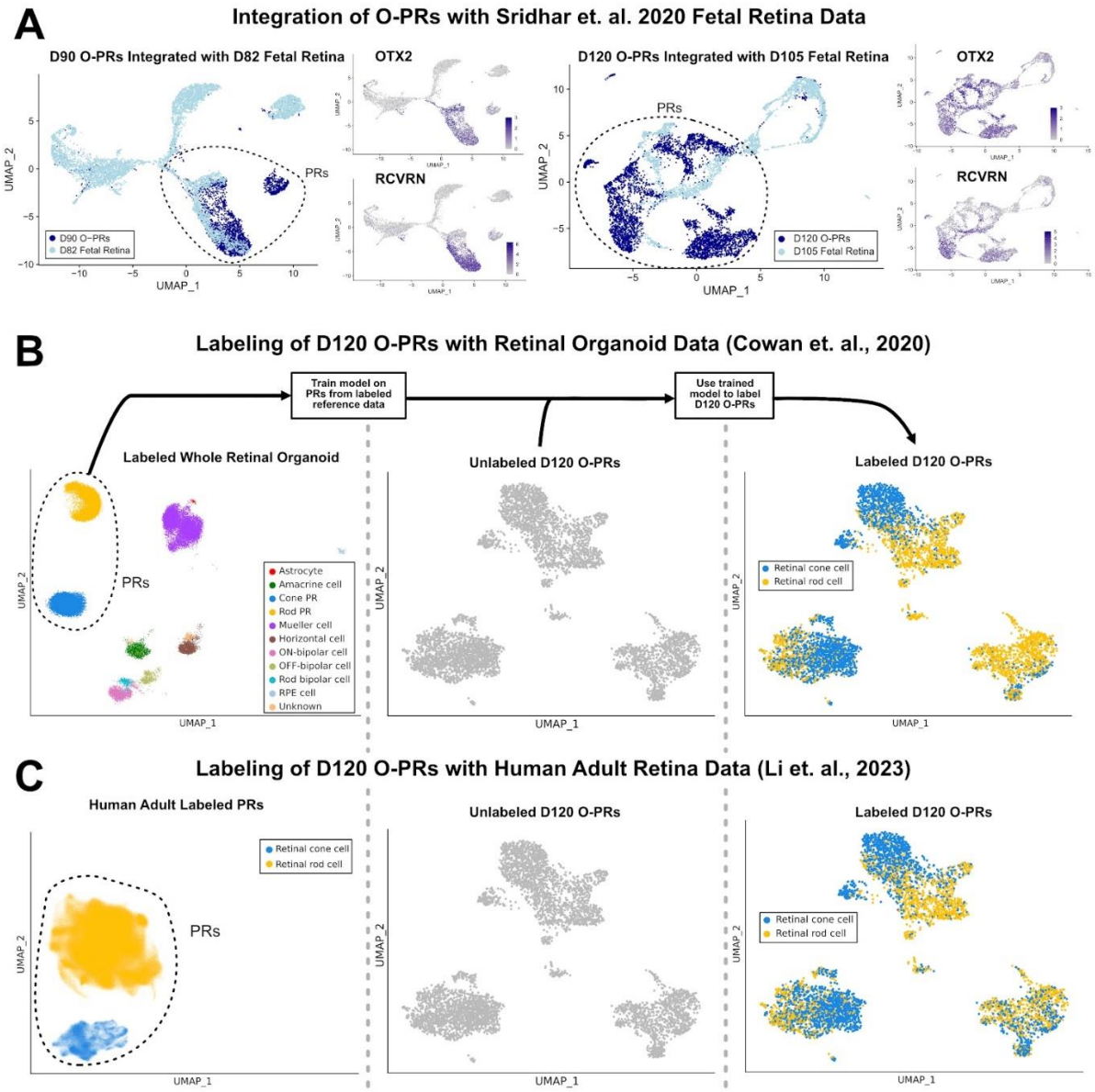


Figure 4: Day 90 and 120 O-PRs compared with fetal, adult and organoid Datasets.

(A) Integration of day 90 (left) and day 120 (right) O-PRs with human fetal retina scRNA-seq data taken from Sridar et. al., 2020. PR population is circled, as determined by expression of markers *OTX2* and *RCVRN*. **(B-C)** Label transfer of day 120 O-PRs using **(B)** retinal organoid data from Cowan et. al., 2020 and **(C)** human adult retina data from Li et. al., 2023 as references. PR populations are circled, determined by labels from reference datasets.

References

1. Pascolini, D. & Mariotti, S. P. Global estimates of visual impairment: 2010. *Br. J. Ophthalmol.* **96**, 614–618 (2012).

2. Benfenati, F. & Lanzani, G. New technologies for developing second generation retinal prostheses. *Lab Anim.* **47**, 71–75 (2018).

3. Meyer, J. G., Garcia, T. Y., Schilling, B., Gibson, B. W. & Lamba, D. A. Proteome and Secretome Dynamics of Human Retinal Pigment Epithelium in Response to Reactive Oxygen Species. *Sci. Rep.* **9**, 15440 (2019).

4. Agarwal, D., Do, H., Mazo, K. W., Chopra, M. & Wahlin, K. J. Restoring vision and rebuilding the retina by Müller glial cell reprogramming. *Stem Cell Res.* **66**, 103006 (2023).

5. Cepko, C. L., Austin, C. P., Yang, X., Alexiades, M. & Ezzeddine, D. Cell fate determination in the vertebrate retina. *Proc. Natl. Acad. Sci.* **93**, 589–595 (1996).

6. Wahlin, K. J., Maruotti, J. A., Sripathi, S. R., Ball, J., Angueyra, J. M., Kim, C., Grebe, R., Li, W., Jones, B. W. & Zack, D. J. Photoreceptor Outer Segment-like Structures in Long-Term 3D Retinas from Human Pluripotent Stem Cells. *Sci. Rep.* **7**, 766 (2017).

7. Phillips, M. J., Capowski, E. E., Petersen, A., Jansen, A. D., Barlow, K., Edwards, K. L. & Gamm, D. M. Generation of a rod-specific NRL reporter line in human pluripotent stem cells. *Sci. Rep.* **8**, 2370 (2018).

8. Bhutto, I. & Lutty, G. Understanding age-related macular degeneration (AMD): Relationships between the photoreceptor/retinal pigment epithelium/Bruch's membrane/choriocapillaris complex. *Mol. Aspects Med.* **33**, 295–317 (2012).

9. Feng, K. M., Tsung, T.-H., Chen, Y.-H. & Lu, D.-W. The Role of Retinal Ganglion Cell Structure and Function in Glaucoma. *Cells* **12**, 2797 (2023).

10. Lowry, W. E., Richter, L., Yachechko, R., Pyle, A. D., Tchieu, J., Sridharan, R., Clark, A. T. & Plath, K. Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl. Acad. Sci.* **105**, 2883–2888 (2008).

11. Chen, P., Hao, W., Rife, L., Wang, X. P., Shen, D., Chen, J., Ogden, T., Van Boemel, G. B., Wu, L., Yang, M. & Fong, H. K. W. A photic visual cycle of rhodopsin regeneration is dependent on Rgr. *Nat. Genet.* **28**, 256–260 (2001).

12. Sakami, S., Kolesnikov, A. V., Kefalov, V. J. & Palczewski, K. P23H opsin knock-in mice reveal a novel step in retinal rod disc morphogenesis. *Hum. Mol. Genet.* **23**, 1723–1741 (2014).

13. Sakamoto, K., McCluskey, M., Wensel, T. G., Naggert, J. K. & Nishina, P. M. New mouse models for recessive retinitis pigmentosa caused by mutations in the Pde6a gene. *Hum. Mol. Genet.* **18**, 178–192 (2009).

14. Jiang, H., Xiong, S. & Xia, X. Retinitis pigmentosa-associated rhodopsin mutant T17M induces endoplasmic reticulum (ER) stress and sensitizes cells to ER stress-induced cell death. *Mol. Med. Rep.* **9**, 1737–1742 (2014).

15. Fletcher, E. L., Jobling, A. I., Vessey, K. A., Luu, C., Guymer, R. H. & Baird, P. N. Chapter 6 - Animal Models of Retinal Disease. in *Progress in Molecular Biology and Translational Science* (eds. Chang, K. T. & Min, K.-T.) vol. 100 211–286 (Academic Press, 2011).

16. 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., Gronostajski, R. M., Fertig, E. J., Goff, L. A. & Blackshaw, S. Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron* **102**, 1111-1126.e5 (2019).

17. Lukowski, S. W., Lo, C. Y., Sharov, A. A., Nguyen, Q., Fang, L., Hung, S. S., Zhu, L., Zhang, T., Grünert, U., Nguyen, T., Senabouth, A., Jabbari, J. S., Welby, E., Sowden, J. C., Waugh, H. S., Mackey, A., Pollock, G., Lamb, T. D., Wang, P., Hewitt, A. W., Gillies, M. C., Powell, J. E. & Wong, R. C. A single-cell transcriptome atlas of the adult human retina. *EMBO J.* **38**, e100811 (2019).

18. Cowan, C. S., Renner, M., De Gennaro, M., Gross-Scherf, B., Goldblum, D., Hou, Y., Munz, M., Rodrigues, T. M., Krol, J., Szikra, T., Cuttat, R., Waldt, A., Papasaikas, P., Diggelmann, R., Patino-Alvarez, C. P., Galliker, P., Spirig, S. E., Pavlinic, D., Gerber-Hollbach, N., Schuierer, S., Srdanovic, A., Balogh, M., Panero, R., Kusnyerik, A., Szabo, A., Stadler, M. B., Orgül, S., Picelli, S., Hasler, P. W., Hierlemann, A., Scholl, H. P. N., Roma, G., Nigsch, F. & Roska, B. Cell Types of the Human Retina and Its Organoids at Single-Cell Resolution. *Cell* **182**, 1623-1640.e34 (2020).

19. Menon, M., Mohammadi, S., Davila-Velderrain, J., Goods, B. A., Cadwell, T. D., Xing, Y., Stemmer-Rachamimov, A., Shalek, A. K., Love, J. C., Kellis, M. & Hafler, B. P. Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat. Commun.* **10**, 4902 (2019).

20. Date, R. C., Jesudasen, S. J. & Weng, C. Y. Applications of Deep Learning and Artificial Intelligence in Retina. *Int. Ophthalmol. Clin.* **59**, 39–57 (2019).

21. Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B. S., Waldstein, S. M. & Bogunović, H. Artificial intelligence in retina. *Prog. Retin. Eye Res.* **67**, 1–29 (2018).

22. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., Liu, Y., Samaran, J., Misrachi, G., Nazaret, A., Clivio, O., Xu, C., Ashuach, T., Gabitto, M., Lotfollahi, M., Svensson, V., da Veiga Beltrame, E., Kleshchevnikov, V., Talavera-López, C., Pachter, L., Theis, F. J., Streets, A., Jordan, M. I., Regier, J. & Yosef, N. A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022).

23. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

24. Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M. I. & Yosef, N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).

25. Enhancing Reproducibility through Rigor and Transparency | grants.nih.gov. https://grants.nih.gov/policy/reproducibility/index.htm.

26. Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A. & Honeyman, T. Introducing the FAIR Principles for research software. *Sci. Data* **9**, 622 (2022).

27. Chopra, M. & Agarwal, D. Human retinal ganglion cell neurons generated by synchronous BMP inhibition and transcription factor mediated reprogramming. Github at https://github.com/WahlinLab/Human_RGC-iN

28. Wahlin, K. J., Cheng, J., Jurlina, S. L., Jones, M. K., Dash, N. R., Ogata, A., Kibria, N., Ray, S., Eldred, K. C., Kim, C., Heng, J. S., Phillips, J., Johnston, R. J., Gamm, D. M., Berlinicke, C. & Zack, D. J. CRISPR Generated SIX6 and POU4F2 Reporters Allow Identification of Brain and Optic Transcriptional Differences in Human PSC-Derived Organoids. *Front. Cell Dev. Biol.* **9**, 764725 (2021).

29. Harkin, J., Peña, K. H., Gomes, C., Hernandez, M., Lavekar, S. S., So, K., Lentsch, K., Feder, E. M., Morrow, S., Huang, K.-C., Tutrow, K. D., Morris, A., Zhang, C. & Meyer, J. S. A highly reproducible and efficient method for retinal organoid differentiation from human pluripotent stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2317285121 (2024).

30. Clark, I. C., Fontanez, K. M., Meltzer, R. H., Xue, Y., Hayford, C., May-Zhang, A., D'Amato, C., Osman, A., Zhang, J. Q., Hettige, P., Ishibashi, J. S. A., Delley, C. L., Weisgerber, D. W., Replogle, J. M., Jost, M., Phong, K. T., Kennedy, V. E., Peretz, C. A. C., Kim, E. A., Song, S., Karlon, W., Weissman, J. S., Smith, C. C., Gartner, Z. J. & Abate, A. R. Microfluidics-free single-cell genomics with templated emulsification. *Nat. Biotechnol.* **41**, 1557–1566 (2023).

31. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

32. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., McElrath, J. M., Blish, C. A., Gottardo, R., Smibert, P. & Satija, R. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

33. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

34. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

35. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

36. Chopra, M. Validation of Human Organoid-Derived PRs. Github at https://github.com/recursivelymanan/Validation-of-Human-Organoid-Derived-PRs

37. Agarwal, D., Dash, N., Mazo, K. W., Chopra, M., Avila, M. P., Patel, A., Wong, R. M., Jia, C., Do, H., Cheng, J., Chiang, C., Jurlina, S. L., Roshan, M., Perry, M. W., Rho, J. M., Broyer, R., Lee, C. D., Weinreb, R. N., Gavrilovici, C., Oesch, N. W., Welsbie, D. S. & Wahlin, K. J. Human retinal ganglion cell neurons generated by synchronous BMP inhibition and transcription factor mediated reprogramming. *Npj Regen. Med.* **8**, 1–18 (2023).

38. Guan, Y., Wang, Y., Zheng, D., Xie, B., Xu, P., Gao, G. & Zhong, X. Generation of an RCVRN-eGFP Reporter hiPSC Line by CRISPR/Cas9 to Monitor Photoreceptor Cell Development and Facilitate the Cell Enrichment for Transplantation. *Front. Cell Dev. Biol.* **10**, 870441 (2022).

39. 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., Glass, I. & Reh, T. A. Single-Cell Transcriptomic Comparison of Human Fetal Retina, hPSC-Derived Retinal Organoids, and Long-Term Retinal Cultures. *Cell Rep.* **30**, 1644-1659.e4 (2020).

40. Xu, Y., Sulaiman, P., Feddersen, R. M., Liu, J., Smith, R. G. & Vardi, N. Retinal ON Bipolar Cells Express a New PCP2 Splice Variant That Accelerates the Light Response. *J. Neurosci.* **28**, 8873–8884 (2008).

41. Zou, C. & Levine, E. M. Vsx2 controls eye organogenesis and retinal progenitor identity via homeodomain and non-homeodomain residues required for high affinity DNA binding. *PLoS Genet.* **8**, e1002924 (2012).

42. Schmitt, S., Aftab, U., Jiang, C., Redenti, S., Klassen, H., Miljan, E., Sinden, J. & Young, M. Molecular characterization of human retinal progenitor cells. *Invest. Ophthalmol. Vis. Sci.* **50**, 5901–5908 (2009).

43. Newton, F. & Megaw, R. Mechanisms of Photoreceptor Death in Retinitis Pigmentosa. *Genes* **11**, 1120 (2020).

44. Li, J., Wang, J., Ibarra, I. L., Cheng, X., Luecken, M. D., Lu, J., Monavarfeshani, A., Yan, W., Zheng, Y., Zuo, Z., Colborn, S. L. Z., Cortez, B. S., Owen, L. A., Tran, N. M., Shekhar, K., Sanes, J. R., Stout, J. T., Chen, S., Li, Y., DeAngelis, M. M., Theis, F. J. & Chen, R. Integrated multi-omics single cell atlas of the human retina. 2023.11.07.566105 Preprint at https://doi.org/10.1101/2023.11.07.566105 (2023).

45. Hendrickson, A., Bumsted-O'Brien, K., Natoli, R., Ramamurthy, V., Possin, D. & Provis, J. Rod photoreceptor differentiation in fetal and infant human retina. *Exp. Eye Res.* **87**, 415–426 (2008).

46. Bumsted O'Brien, K. M., Cheng, H., Jiang, Y., Schulte, D., Swaroop, A. & Hendrickson, A. E. Expression of photoreceptor-specific nuclear receptor NR2E3 in rod photoreceptors of fetal human retina. *Invest. Ophthalmol. Vis. Sci.* **45**, 2807–2812 (2004).

47. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A. & Trapnell, C. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

48. Cepko, C. Intrinsically different retinal progenitor cells produce specific types of progeny. *Nat. Rev. Neurosci.* **15**, 615–627 (2014).

49. Austin, C. P., Feldman, D. E., Ida, J. A. & Cepko, C. L. Vertebrate retinal ganglion cells are selected from competent progenitors by the action of Notch. *Dev. Camb. Engl.* **121**, 3637–3650 (1995).

50. Reh, T. A. & Kljavin, I. J. Age of differentiation determines rat retinal germinal cell phenotype: induction of differentiation by dissociation. *J. Neurosci. Off. J. Soc. Neurosci.* **9**, 4179–4189 (1989).

51. Gomes, F. L. A. F., Zhang, G., Carbonell, F., Correa, J. A., Harris, W. A., Simons, B. D. & Cayouette, M. Reconstruction of rat retinal progenitor cell lineages in vitro reveals a surprising degree of stochasticity in cell fate decisions. *Dev. Camb. Engl.* **138**, 227–235 (2011).

52. Cayouette, M., Barres, B. A. & Raff, M. Importance of intrinsic mechanisms in cell fate decisions in the developing rat retina. *Neuron* **40**, 897–904 (2003).

53. Bernardos, R. L., Barthel, L. K., Meyers, J. R. & Raymond, P. A. Late-stage neuronal progenitors in the retina are radial Müller glia that function as retinal stem cells. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 7028–7040 (2007).

54. Ooto, S., Akagi, T., Kageyama, R., Akita, J., Mandai, M., Honda, Y. & Takahashi, M. Potential for neural regeneration after neurotoxic injury in the adult mammalian retina. *Proc. Natl. Acad. Sci.* **101**, 13654–13659 (2004).

55. Karl, M. O., Hayes, S., Nelson, B. R., Tan, K., Buckingham, B. & Reh, T. A. Stimulation of neural regeneration in the mouse retina. *Proc. Natl. Acad. Sci.* **105**, 19508–19513 (2008).

56. Fischer, A. J. & Reh, T. A. Müller glia are a potential source of neural regeneration in the postnatal chicken retina. *Nat. Neurosci.* **4**, 247–252 (2001).

57. Fausett, B. V. & Goldman, D. A Role for α1 Tubulin-Expressing Müller Glia in Regeneration of the Injured Zebrafish Retina. *J. Neurosci.* **26**, 6303–6313 (2006).