bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Generation of human neural retina transcriptome atlas by single cell RNA sequencing

Samuel W. Lukowski[1,2,#], Camden Y. Lo[3,#], Alexei Sharov[4], Quan H. Nguyen[1], Lyujie Fang[5,6,7], Sandy S.C. Hung[5,6], Ling Zhu[8], Ting Zhang[8], Tu Nguyen[5,6], Anne Senabouth[1], Jafar S. Jabbari[9], Emily Welby[10], Jane C. Sowden[10], Hayley S. Waugh[11], Adrienne Mackey[11], Graeme Pollock[11], Trevor D. Lamb[12], Peng-Yuan Wang[13,14], Alex W. Hewitt[5,6,15], Mark Gillies[8], Joseph E. Powell[1,16,^], Raymond C.B. Wong[5,6,17,^]

[1] Institute for Molecular Bioscience, University of Queensland, Australia
[2] The University of Queensland Diamantina Institute, Translational Research Institute, Australia
[3] Private Consultancy
[4] National Institute for Aging, National Institutes of Health, United States
[5] Centre for Eye Research Australia, Australia
[6] Ophthalmology, Department of Surgery, University of Melbourne, Australia
[7] Jinan University, China
[8] Save Sight Institute, University of Sydney, Australia
[9] Australian Genome Research Facility, Australia
[10] Stem Cells and Regenerative Medicine Section, NIHR Great Ormond Street Hospital Biomedical Research Centre, UCL Great Ormond Street Institute of Child Health, UK
[11] Lions Eye Donation Services, Australia
[12] John Curtin School of Medical Research, The Australian National University, Australia
[13] Department of Chemistry and Biotechnology, Swinburne University of Technology, Australia
[14] Center for Human Tissue and organs Degeneration, Institute of Biomedicine and Biotechnology, Shenzhen Institute of Advanced Technology, Chinese Academy of Science, China
[15] Menzies Institute for Medical Research, University of Tasmania, Australia
[16] Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Australia
[17] Shenzhen Eye Hospital, China

[#] equal first-authors.
[^] equal senior-authors

Lead contact/correspondence: Dr Raymond Wong, Centre for Eye Research Australia, Level 6, 75 Commercial Road, Melbourne, VIC 3004, Australia; Phone: +613 85321962; email: wongcb@unimelb.edu.au

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Summary

The retina is a highly specialized neural tissue that senses light and initiates image processing. Although the functional organisation of specific cells within the retina has been well-studied, the molecular profile of many cell types remains unclear in humans. To comprehensively profile cell types in the human retina, we performed single cell RNA-sequencing on 20,009 cells obtained post-mortem from three donors and compiled a reference transcriptome atlas. Using unsupervised clustering analysis, we identified 18 transcriptionally distinct clusters representing all known retinal cells: rod photoreceptors, cone photoreceptors, Müller glia cells, bipolar cells, amacrine cells, retinal ganglion cells, horizontal cells, retinal astrocytes and microglia. Notably, our data reveal novel subpopulations of rod photoreceptors that can be distinguished by *MALAT1* expression levels. We also demonstrated the use of this retina transcriptome atlas to benchmark pluripotent stem cell-derived cone photoreceptors and an adult Müller glia cell line. This work provides an important reference with unprecedented insights into the transcriptional landscape of human retinal cells, which is fundamental to our understanding of retinal biology and disease.

Keywords

Retina, transcriptome, single cell RNA sequencing, photoreceptor subtypes

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

The eye is a highly specialised sensory organ in the human body. Sight is initiated by the conversion of light into an electrical signal in the photoreceptors of the neurosensory retina. The rod photoreceptors are responsible for light detection at extremely low luminance, while the cone photoreceptors are responsible for colour detection and operate at moderate and higher levels. Following preprocessing, by horizontal, bipolar and amacrine cells, the resultant signal is transferred via ganglion cells to the brain. Neurotransmitter support is provided by Müller glia, retinal astrocytes and microglial cells. Inherited retinal diseases are becoming the leading cause of blindness in working age adults, with loci in over 200 genes associated with retinal diseases (RetNet: https://sph.uth.edu/retnet/), often involving specific retinal cell types. Knowledge of the transcriptome profile of individual retinal cell types in humans is important to understand the cellular diversity in the retina, as well as the study of retinal genes that contribute to disease in individual retinal cell types. [1–5]

The transcriptome profiles of whole human retina from adults [6–10] and during fetal development [11,12] have been previously described. However, these studies only assayed the averaged transcriptional signatures across all cell types, meaning that the cellular heterogeneity in the retina is lost. As such, the transcriptional pathways that underlie the highly specialised function of many human retinal cell types remain unclear; including the rod and cone photoreceptors, Müller glia cells, horizontal cells, and amacrine cells. Recent advances in RNA sequencing and microfluidic platforms have dramatically improved the accessibility of single cell transcriptomics, with increased throughput at a lower cost. Critically, single-cell microfluidics and low-abundance RNA library chemistries allow accurate profiling of the transcriptome of individual cell types. This has been demonstrated in the mouse, where transcriptome profiles of the mouse retina [13] and retinal bipolar cells [14] have been described at the single cell level using the Drop-seq method [13]. These studies provided a molecular classification of the mouse retina and identified novel markers for specific cell types, as well as novel candidate cell types in the retina. Recently, single cell transcriptomics was used to analyse the human retina. Phillips et al. have profiled a total of 139 adult retina cells using the C1 Fluidigm platform [15], but the limited number of profiled cells presents challenges in the annotation and accurate identification of individual retinal cell types. Moreover, a flow cytometry approach was used to isolate 65 human fetal cone photoreceptors followed by scRNA-seq profiling [16].

Herein we report the generation of a human neural retina transcriptome atlas using 20,009 single cells from three donors. Our data provide new insights into the transcriptome profile of major human retinal cell types and establish a high cellular-resolution reference of the human neural retina, which will have implications for identification of biomarkers and understanding retinal cell biology and diseases.

Results

Preparation of human neural retinal samples and generation of single cell transcriptome atlas

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We obtained post-mortem human adult eyes approved for research purposes following corneal transplantation. As the transcriptome profile of human retina pigment epithelium cells has already been reported [17,18], we focused solely on the neural retina layers. We extracted the neural retina from eight donor eyes. We observed consistent cell viability across retinal tissues retrieved within 15 hours post-mortem (Supplementary figure 1A) and found that donor age does not impact negatively on cell viability in the extracted neural retina (Supplementary figure 1B). To minimize potential risk of mRNA degradation due to reduced cell viability, we selected three donor samples retrieved within 15 hours post-mortem (Supplementary Table 1) and analysed them with single cell RNA sequencing (scRNA-seq) using the 10X Genomics Chromium platform.

Sequence data from five scRNA-seq libraries derived from the three neural retinal samples were pooled for processing and analysis. From 23,000 cells, we obtained an average of 40,232 reads per cell and 1,665 UMIs (unique transcripts) per cell. Following quality control and filtering in Seurat, our final dataset contained 20,009 cells, which were taken forward for further analysis.

The scRNA-seq data was initially analysed using an unsupervised graph clustering approach implemented in Seurat (version 2.2.1) to classify individual cells into cell subpopulations according to similarities in their transcriptome profiles. Overall, the cells were classified into 18 transcriptionally distinct clusters (Figure 1A, Supplementary figure 2). We first assessed the variation between donor samples and between library preparations (Supplementary table 2, supplementary figure 3 and 4). Interestingly, although many of the identified clusters are present in all three donor retinal samples, we also observed several donor-specific clusters corresponding to rod photoreceptors (Figure 1B, Supplementary figure 3A and 3B). In contrast, we observed minimal variation between two different libraries prepared from the same donor sample, supporting the quality of the scRNA-seq datasets in this study (Supplementary figure 4).

Identification of major cell types in the human retina using scRNA-seq

Based on known markers [13,14,19–24], we were able to assign cell identities to the 16 of the 18 clusters (Figure 1A, 1C, 1E), corresponding to rod photoreceptors (*PDE6A, CNGA1, RHO*), cone photoreceptors (*ARR3, GNGT2, GUCA1C*), Müller glia (*RLBP1/CRALBP*), retinal astrocytes (*GFAP*), microglia (*HLA-DPA1, HLA-DPB1, HLA-DRA*), bipolar cells (*VSX2, OTX2*), retinal ganglion cells (*NEFL, GAP43, SNCG*), amacrine cells (*GAD1, CALB1, CHAT*) and horizontal cells (*ONECUT1, ONECUT2*). The expression of selected marker genes are displayed in *t*-SNE plots (Figure 1E). Two clusters (C5 and C14) express markers from multiple retinal cell types (Supplementary figure 5), thus we were unable to assign cell identities to these 2 clusters and they were excluded from further analysis. Interestingly, our data demonstrated multiple transcriptionally distinct clusters within the rod photoreceptors (6 clusters) and bipolar cells (3 clusters). In contrast, only one cluster was detected for cone photoreceptors, Müller glia, retinal

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ganglion cells, horizontal cells, amacrine cells, retinal astrocytes and microglia respectively. Correlation analysis confirmed the similarity between clusters within the same cell type (Figure 1F). As expected, we observed high correlations between the expression levels of transcripts within photoreceptor cell types (rod and cones), as well as glial cells (retinal astrocytes and Müller glia) and other retinal neurons (bipolar cells, retinal ganglion cells, amacrine cells and horizontal cells). The composition of cell populations across our three donors show that the majority of the cells in human neural retina were rod photoreceptors (~74%), followed by bipolar cells (~10%) and Müller glia (~3%) (Figure 1D). These results are similar to those reported in mice, where rod photoreceptors and bipolar cells form the majority of cells in the retina [13,25].

To identify genes whose expression was specific to a given cell type, we performed differential gene expression analysis to identify marker genes for each cluster (Figure 1E). We subsequently extracted membrane-related proteins from gene ontology annotations to identify surface markers, which can be used to develop immuno-based methods to isolate primary culture of individual retinal cell types. Figure 1H lists the identified surface markers for individual retinal cell types. We also assessed the gene expression of a panel of commonly known markers in amacrine cells and bipolar cells (Supplementary figure 6-7), as well as a panel of markers for subtype identification recently identified in mouse scRNA-seq studies [13,14]. In summary, we profiled the transcriptomes of all major cell types in the human retina in the presented dataset. Due to their abundance, for the subsequent analyses we focused on the photoreceptors and glial cells.

Presence of heterogeneous rod photoreceptor subpopulations in the human retina

We profiled 14,759 rod photoreceptors and showed that they can be classified into six populations with distinct gene expressions (c0, c1, c2, c3, c4, c7; Supplementary table 3). We assessed these six clusters with a panel of 7 known rod or pan-photoreceptor markers (Figure 2A). Our results suggest differential expression patterns among the 7 markers. All 7 rod markers are highly abundant, consistent with previous scRNA-seq studies of mouse and human retina [13,15]. The 7 markers showed differential expression patterns in the 6 identified rod photoreceptor clusters. In particular, *RHO*, *GNGT1* and *SAG* have the highest levels of rod marker detected, followed by *NRL*, *ROM1*, *GNAT1* and *CNGA1*. We also noted that *ROM1* is expressed in both rod and cone photoreceptors, which is consistent with previous studies [26]. Importantly, many rod photoreceptor clusters consist of a majority of cells from a single donor (>90% for c0, c2, c4 and >80% for c1, c7; Figure 2B). It is possible that this observation is due to the systematic biases such as differences in tissue retrieval time, age of donors, or other sample preparation variation. The exception is cluster c3, which is well represented by all three donors.

Taking into account donor variations, we set out to further define and classify the difference between rod photoreceptor clusters. We observed that *MALAT1*, a long non-coding RNA that plays a role in retinal maintenance and diseases [27], was robustly expressed in ~66% of the

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

identified rod photoreceptors (9,750 cells) whilst the rest had little to no expression (5,009 cells; Figure 2C, Supplementary figure 8A). As such, we utilized *MALAT1* expression as a discriminator and investigated differences between rod photoreceptors with high expression (*MALAT1-hi*; > 4.68 normalised transcripts per cell) or low expression (*MALAT1-lo*; < 4.68 normalised transcripts per cell). *MALAT1-hi* and *-lo* rod photoreceptors were consistently found in each donor and library samples, with *MALAT1-hi* accounting for ~66%, 90% and 36% of the rods in donors #1, #2 and #3 respectively (Figure 2D). To further validate this finding, we performed RNA *in situ* hybridization in another three donor retinal samples. We consistently observed the presence of *MALAT1-hi* and *-lo* subpopulations of rod photoreceptors in all retinal samples (Figure 2E, Supplementary figure 9). Together, these results showed the presence of heterogeneity within rod photoreceptors that can be distinguished by *MALAT1* expression .

We pooled all rod photoreceptors together (Clusters c0-4, c7) and performed *t*-SNE and differential gene expression assessment between *MALAT1-hi* and *-lo* rod photoreceptors. Stochastic embedding displayed *MALAT1* expression as conforming well to cell-to-cell neighborhood clustering and delineation (Supplementary figure 8B), with *MALAT1-hi* cells up-regulated for nuclear components and transcription factors such as *AHI1* which is important for photoreceptor outer segment development [28], *DDB1* which has been described to play a role in eye development [29], and *DDX17* which is a member of the DEAD box protein family of RNA helicases that function as transcriptional co-regulator [30]; Figure 2E). In addition, the *MALAT1-hi* cells also showed increased expression of *PDE6A*, *RP1* and *CNGB1*, which are the genetic causes of retinitis pigmentosa when mutated [31]. Further analysis indicated that many genes involved in the phototransduction pathways are enriched in *MALAT1-hi* rod photoreceptors (Figure 2F). On the other hand, *MALAT1-lo* cells showed up-regulation of the transcription factor *NRL* which is essential for rod photoreceptor specification [32], as well as mitochondrial genes (e.g. *MT-CO3, MT-ND4, MT-ND3, MT-CYB*), ribosomal genes (e.g. *RPL36, RPL39, RPL143*) and other phototransduction genes (*PDE6G, GNGT1*; Figure 2E). Taken together our results suggest the presence of two rod photoreceptor subpopulations in human retina that can be distinguished by *MALAT1* expression levels.

Transcriptome profile of cone subtypes in the human retina

We detected 564 cone photoreceptor cells in our scRNA-seq data, which are distinguishable from the other cell types by the expression of the cone marker genes *ARR3, CNBG3, GNAT2, GNGT2, GRK7, GUCA1C, PDE6C, PDE6H, OPN1LW, RXRG* and *THRB* (Figure 3A). All 11 marker genes analysed show specific expression patterns in the cone cluster (C10). We set out to further assess the composition of the cone cluster. In humans, there are three identified subtypes of cone photoreceptor which can be distinguished by expression of a sole opsin gene: *OPN1SW*-positive S-cones, *OPN1MW*-positive M-cones and *OPN1LW*-positive L-cones respond preferentially to shorter, medium and longer wavelengths responsible for colour vision [33].

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Notably *OPN1LW* and *OPN1MW* exhibit ~98% sequence homology and are unable to be distinguished by 3' sequencing utilised in this study. By quantifying the number of cells that express the opsin genes, our results showed that the majority of the cone cluster are L/M cones (73.22%) and S-cones in much lower proportion (3.19%, Figure 3B), at levels consistent with those estimated by a previous adaptive optics and photo-bleaching study [33]. As expected, the identified cone photoreceptors only express either *OPN1SW* or *OPN1LW/MW* (Figure 3C).

To further study the molecular differences and identify molecular markers for the cone subtypes, we performed differential gene expression analysis to determine genes that can distinguish the cone subtypes. Our results identified a panel of genes that differentially marked S-cones (e.g. *CCDC136*, *RAMP1*, *LY75*, *CADM3*, *TFPI2*, *CRHBP*, *RAB17*, *UPB1*, *RRAD*, *SLC12A1*) and L/M-cones (e.g. *THRB*, *KIF2A*, *LBH*, *PGP*, *CHRNA3*, *AHI1*, *LIMA1*; Figure 3D). Together these results detailed the molecular profiles and identified marker genes that can distinguish the cone subtypes in human.

Assessment of glial cells in human retina

Next, we focused on two related glial cell types in the human retina, the Müller glia and the retinal astrocytes. Our scRNA-seq data has profiled a total of 2,723 Müller glia cells which classified into a single cluster (C9) and 49 retinal astrocytes which form a single cluster (C16). Figure 3E shows the expression of a panel of 9 commonly used markers for Müller glia and retinal astrocytes. Our results demonstrated that many of these markers are present in both Müller glia and retinal astrocytes at differential expression levels. *VIM*, *GLUL* and *S100A1* marked both Müller glia and retinal astrocytes at high expression levels. *GFAP* represents a reliable marker for retinal astrocytes, and its expression is consistent with a previous report [20]. Notably, Müller glia are low in *GFAP* expression, indicating they are not in an activated state commonly caused by stresses and reactive gliosis [34]. The *S100B* is also expressed in retinal astrocytes at varying levels but absent in Müller glia. Conversely, Müller glia can be distinguished from retinal astrocytes by high expression levels of *RLBP1*, and *RGR* to a lesser extent. Together these results provide insights into the differential expression patterns of known glial markers in Müller glia and retinal astrocytes in human.

As glial cell proliferation has been linked to a range of pathological conditions including retinal gliosis and retinal injury [35], this provides a means of assessing the health of the profiled retinas. We assigned a cell cycle phase score to each cell using gene expression signatures for the G1, S, G2, and G2/M phases [36]; Supplementary figure 10). We determined that most of the Müller glial cells expressed genes indicative of cells in G1 phase (75%), suggesting they are not proliferative. These results demonstrated the absence of hallmarks of gliosis/retinal injury and support the quality of the donor retinas profiled.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Using the human neural retina transcriptome atlas for benchmarking

To demonstrate the use of our dataset as a benchmarking reference, we compared the scRNA-seq profiles of distinct cell types generated using alternative methods, including fetal human cone photoreceptors, human induced pluripotent stem cell derived-cone photoreceptors (hiPSC-cone; [16], and a sample of adult human retina with 139 cells [15]. Correlation analysis demonstrated that the adult human retina sample showed highest similarity to rod photoreceptor (0.63, Supplementary figure 11), which is expected as rod photoreceptors represent the majority of the cells in the retina. Interestingly, our results also showed that the transcriptome of hiPSC-cone after 15 weeks of differentiation exhibited the highest similarity to cone photoreceptors, and low similarities to all other retinal cell types (Figure 4A, Supplementary figure 11). In particular, hiPSC-cone showed high similarities to fetal cone photoreceptors and adult cone photoreceptors (0.71 and 0.61 respectively), and a much lower similarity to adult rod photoreceptors (0.33). In support of this, principal component analysis demonstrated that the hiPSC-cone are closer to fetal cone photoreceptors, rather than the adult counterpart (Figure 4B, Supplementary figure 11). These results confirmed direct differentiation of hiPSCs to cone photoreceptors with good quality, and the hiPSC-derived cone photoreceptors are closer to fetal origin compared to their adult counterpart.

In another benchmarking example, we set out to assess the potential differences between *in vitro* cell lines compared to adult cells *in vivo*. In this regard, we compared the spontaneously immortalised human Müller glia cell line MIO-M1 [37] to all the retinal cell types identified in our dataset. Using scRNA-seq, we profiled 7,150 MIO-M1 cells with 23,987 reads per cell post-normalization corresponding to 3,421 detected genes. t-SNE analysis showed that the MIO-M1 cells formed one cluster that is transcriptionally distinct from all retina cell types identified in the human neural retina dataset (Fig 4C, Supplementary figure 11). Correlation analysis showed that MIO-M1 displayed similarities to retinal glial cells, with higher similarity to astrocytes compared to Müller glia (0.63 and 0.46 respectively, Fig 4D). In particular, we identified that MIO-M1 express high levels of thymosin beta 4 gene (*TMSB4X*), which has been linked to glioma malignancy [38], as well as the calcyclin gene (*S100A6*), which is implicated in macular or cone associated diseases [39]; Figure 4E). Together, our results highlighted the similarities and differences of MIO-M1 to adult retinal glial cells in human.

Discussion

The data presented here describe the generation of a detailed reference transcriptome atlas of the human neural retina at the single-cell level. The establishment of reference transcriptome maps for individual cell types in the retina provide unprecedented insights into the signals that define retinal cell identity and advance our understanding of the retina. This human neural retina transcriptome data can be used as a benchmark to assess the quality and maturity of pluripotent

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

stem cell-derived retinal cells, such as retinal ganglion cells [40–42] and photoreceptors [43]. We also confirmed that a relatively low level of sequencing depth (~40,232 reads per cell) is sufficient for identification and classification of cell types in a complex tissue like retina.

One of the most interesting observations is the presence of heterogeneous subpopulations within known retinal cell types. This highlights the sensitivity of using a scRNA-seq approach to capture and classify retinal cell types in an unbiased manner. In particular, our results demonstrated the presence of two subpopulations of rod photoreceptor that display differential expression of *MALAT1*, a finding that has not previously been reported in human or any other species. Previous studies have demonstrated a role of *MALAT1* in regulating the survival of retinal ganglion cells [44] and in pathogenesis of retinal pigment epithelium cells [45]. Our study highlighted that with newly achieved sequencing depth and resolution, heterogeneity can be observed in what were previously thought to be homogeneous cell populations. Indeed, previous studies support the existence of two subpopulations of rod photoreceptors in mouse. Tsukamoto *et al.* demonstrated that a rod subpopulation forms synaptic contacts with cone bipolar cells, whereas another subpopulation forms synapses with rod bipolar cells [46]. More recently, Kim *et al.* showed the presence of two subpopulations in developing mouse rod photoreceptors, with one subpopulation derived from S-cones during development [47]. A key factor in regulating cone-to-rod transition is the MAF family transcription factor *NRL*, where its presence is essential for rod specification and its absence lead to cone specification during development [32]. Interestingly, our data showed that *NRL* expression is significantly higher in *MALAT1-lo* compared to *MALAT1-hi* rod subpopulations in human, which may indicate similarity to the two rod subpopulations found in mouse. Our results also indicated that genes related to phototransduction pathways are differentially expressed in *MALAT1-lo* and *-hi* rod photoreceptors, potentially suggesting functional differences of light responses in these two subpopulations. It is also possible that these two *MALAT1* subpopulations reflect different cellular states rather than functionally distinct subpopulations. Future studies are warranted to investigate the role of *MALAT1* in photoreceptors. Moreover, our transcriptome data revealed rod photoreceptor clusters specific to particular donor retinas, which could be due to the process of sample preparation or donor variations. Further studies with a larger number of donor samples will allow testing of the feasibility of using scRNA-seq to comprehensively analyse the retina in different individuals, such as assessment of the effects of aging or degenerative retinal diseases across different patients.

Another outcome of this study is the assessment of biomarkers that allow identification of major retinal cell types and subtypes. Our results provide new insights into the cone photoreceptor subtypes in human. The cone subtypes are traditionally categorized based on expression of different opsins that allowed for cellular detection of light at various wavelength. While the S-cones are structurally different from the other two cone subtypes, the L-cones and M-cones are structurally similar and difficult to distinguish from each other, except for the opsin they

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

expressed [48]. We report the first description of the transcriptome profiles of S-cones in adult human and highlight novel marker genes that can be used to distinguish them. We also identified the transcriptome and novel marker genes for L/M-cones, however given the high sequence homology, particularly at the 3' end, of *OPN1MW* and *OPN1LW*, we could not confidently separate L-cones and M-cones. In addition, we show that many of the known Müller glia markers are often expressed in retinal astrocytes, and we also provide a detailed assessment of commonly used retinal glial markers showing the differential expression pattern between Müller glia and retinal astrocytes.

Finally, our results highlighted the use of this neural retina transcriptome atlas to benchmark retinal cells derived from stem cells or primary cultures. A major goal of pluripotent stem cell research is to derive cells that are similar to those in adults *in vivo*, which is important for development of stem cell disease models and regenerative medicine [49]. Our analysis shows that hiPSC-derived cone photoreceptors are highly similar to both fetal and adult cones in comparison with all other major retinal cell types. We show that hiPSC-derived cells are more fetal-like than adult-like, which is consistent with other studies [50,51]. We also benchmark a commonly used Müller glia cell line MIO-M1. Our results showed that while this cell line exhibits similarities to adult retinal glial cells, there are also some differences between MIO-M1 and adult Müller glia, such as high expression of the glioma-related gene thymosin beta4 (*TMSB4X)* in MIO-M1. Previous reports have also described differences in gene expression in MIO-M1 to Müller glia, and showed that MIO-M1 express markers for post-mitotic retinal neurons and neural stem cells [37,52]. These results highlighted the potential effects of prolonged *in vitro* culture of primary retinal cells. Collectively, we showed that the human neural retina transcriptome atlas provides an important benchmarking resource to assess the quality of derived retinal cells, which would have implications for stem cell and neuroscience research.

One of the limitations of this study is the finite number of profiled cell types less frequently represented in the retina such as the amacrine cells and the retinal ganglion cells, which are known to be highly complex. The presented dataset is limited in power to accurately identify differences in the transcriptomes of the subtypes in amacrine and retinal ganglion cells. With the identification of surface markers for these cell types in this study, this work lays the foundation for future research using selection and enrichment [14] of these and other cell types (such as horizontal cells or retinal microglia) to improve the resolution of the human neural retina transcriptome atlas. Another limitation is the use of 3' gene expression profiling, which presents a challenge for distinguishing L-cones and M-cones. Given the high sequence homology of *OPN1LW* and *OPN1MW* (98%), future studies using full-length mRNA sequencing of single cone photoreceptor cells would provide greater distinction and classification accuracy of the *OPN1MW* and *OPN1LW*-positive cells.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This study describes the transcriptome of human neural retina at a single cell level and is the first to identify the transcriptome of all major human retinal cell types. Our findings shed light on the molecular differences between subpopulations within the rod photoreceptors and the cone photoreceptors. The presented dataset provides an important roadmap to define the genetic signals in major cell types in the human retina and can be used as a benchmark to assess the quality of stem cell-derived cells or primary retinal cells.

Online Methods

Human retina collection

Collection of patient samples was approved by the Human Research Ethics committee of the Royal Victorian Eye and Ear Hospital (HREC13/1151H) and Save Sight Institute (16/282) and carried out in accordance with the approved guidelines. For scRNA-seq, post-mortem eye globes were collected by the Lions Eye Donation Service (Royal Victorian Eye and Ear Hospital) for donor cornea transplantation. The remaining eye globes were used for dissection to extract the neural retina. The lens, iris and vitreous were removed and the choroid/RPE layers were excluded from the sample collection. Following extraction, the neural retinal samples were dissociated and processed for scRNA-seq right away. Neural retina samples were dissociated into single cells in dissociation solution (2mg/ml papain, 120 Units/ml DNase I) for 15 minutes. The dissociated neural retina was filtered to ensure single cell suspension using a 30µm MACS Smart Strainer (Miltenyi). For sample from Patient SC, the Dead Cell Removal kit (Miltenyi) was utilised to remove dead cells prior to scRNA-seq. However, in our hands we found that the Dead Cell Removal kit only had a modest improvement in the cell viability (~8% improvement, data not shown).

Single cell RNA sequencing (scRNA-seq)

Single cells from three independent neural retina samples were captured in five batches using the 10X Chromium system (10X Genomics). The cells were partitioned into gel bead in emulsions and barcoded cDNA libraries, then prepared using the single cell 3' mRNA kit (V2; 10X Genomics). Single cell libraries were sequenced in 100bp paired-end configuration using an Illumina Hi-Seq 2500 at the Australian Genome Research Facility.

Bioinformatics processing

The 10X Genomics *cellranger* pipeline (version 2.1.0; [53] was used to generate fastq files from raw Illumina BCL files (*mkfastq*) and aligned to the human genome reference GRCh38 using the included STAR alignment software [54]. Next, *cellranger count* was used to generate read count matrices. To overcome the stringent threshold implemented in cellranger that discards real cells under certain conditions, such as populations of cells with a low RNA content, the *--force-cells* parameter was set to 3000 for the donor 1 library and 5000 for donor 2 and 3 libraries. Using the barcode rank plot, these parameters were selected to increase the number of detected cells for

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

further analysis. The *cellranger* aggregation function (*aggr*) was used to combine the 5 libraries and normalize the between-sample library size differences.

Data were imported into the Seurat single cell analysis software (v2.0.1; https://github.com/satijalab/seurat). Quality control of sequenced libraries was performed to remove outlier cells and genes. Cells with 200-2500 detected genes and expressing < 10% mitochondrial genes were retained. Genes were retained in the data if they were expressed in ≥ 3 cells. Additional cell-cell normalization was performed using the LogNormalize method, and inherent variation caused by mitochondrial gene expression and the number of unique molecular identifiers (UMIs) per cell was regressed out.

Clustering was performed on PCA-reduced expression data using the top 20 principal components using the graph-based shared nearest neighbour method (SNN) which calculates the neighborhood overlap (Jaccard index) between every cell and its nearest neighbors.

Prediction of the cell cycle phase of individual retinal cells was performed in Seurat using cell cycle-specific expression data [36]. Briefly, genetic markers associated with G2/M and S phase were used to assign cell scores, and cells expressing neither of the G2/M or S phase markers were classified as being in G1 phase.

Sequencing data for fetal (scRNA-seq) and hiPSC-derived cone photoreceptors (bulk RNA-seq) was obtained from ArrayExpress using the accession numbers E-MTAB-6057 and E-MTAB-6058 [16]. Gene expression matrices were generated from the fastq files using the STAR aligner software. scRNA-seq data from 72 cells were quality-controlled, filtered and then normalised with the scran algorithm [55] as described [16], using the *ascend* (https://github.com/IMB-Computational-Genomics-Lab/ascend) package in R, which resulted in 63 high quality single cell transcriptomes. Bulk RNA-seq data generated from 6 hiPSC-derived cone photoreceptor cultures was filtered such that each gene was represented by at least 10 counts in all samples and normalisation was performed in edgeR using the trimmed mean of M method [56]. Pre-processed scRNA-seq data generated from adult retina [15] was obtained from the Gene Expression Omnibus (GSE98556).

Identification of retinal cell types

Cell types were classified using differential expression analysis, which compared each cluster to all others combined using the Wilcoxon method in Seurat to identify cluster-specific marker genes. Each retained marker gene was expressed in a minimum of 25% of cells and at a minimum $\log_2$ fold change threshold of 0.25.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In paired cluster analyses, differentially expressed genes were considered significant if the adjusted p-value was less than 0.01 (Benjamini-Hochberg correction for multiple testing) and the absolute $\log_2$ expression fold change was ≥ 0.5.

Mapping cells between subpopulations in different samples

To compare subpopulations identified in the merged dataset (5 samples), we applied scGPS (single cell Global Projection between Subpopulations), a machine learning procedure to select optimal gene predictors and to build prediction models that can estimate between-subpopulation transition scores. The transition scores are the probability of cells from one subpopulation that are in the same class with the cells in the other subpopulation. The scores, therefore, estimate the similarity between any two subpopulations. Here, we compared three main subpopulations from sample Retina 2A with all subpopulations in the sample Retina 2B. The source code of the scGPS method is available with open-access (https://github.com/IMB-Computational-Genomics-Lab/scGPS).

Correlation of scRNA-seq data with retinal cell types

The mean expression levels of cells in each cluster were calculated and used to calculate Pearson's correlations in a pairwise manner with each of the other clusters and results were deemed significant if the correlation P-value was less than 0.01.

Pathway analysis

Gene enrichment analysis was performed using Enrichr [57]. The combined score, computed by taking the log of p-value from the Fisher exact test and multiplying by the z-score of the deviation of the expected rank, was used to determine the enrichment ranking for pathways, othologies, transcription factor network and protein network analysis.

Fluorescent in situ hybridization

Donor retinas were first dissected from the eye cup. The retinal tissues were subjected to 30% sucrose cryoprotection and were then frozen in -80°C. Sections were cut on a cryostat (Leica CM3050S) and mounted on glass slides (SuperFrostPlus). The retinal samples were fixed in 3.7% (vol/vol) formaldehyde and hybridized with Stellaris RNA FISH probes (Biosearch Technologies) against *MALAT1* labeled with Quasar 570, following the manufacturer's instructions. Briefly, samples were incubated with Quasar 570-labeled probes at 125nM in hybridization buffer and hybridized 5 hours at 37°C, followed by nuclear counterstain using DAPI. The samples are imaged using a ZEISS confocal laser-scanning microscope (ZEISS, LSM700).

Author Contributions

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

HW, AM coordinated and collected the human donor retinas, GP provided the funding for human donor tissue collection, LF, TN, JJ, SH, RW, LZ, TZ conducted the experiments; SL, CL, AS, RW, QN, EW, JS, TL, LZ, TZ processed and/or analysed the data; PYW, AH, JP, MG and RW contributed to experimental design and data analysis; SL, AH, JP and RW wrote the manuscript.

Acknowledgements

The authors want to thank David Mackey and Holly Chinnery for critical discussion and helpful feedbacks for this study, and Prema Finn for technical assistance for the collection of human donor samples. This work was supported by funding from the Ophthalmic Research Institute of Australia (RW, SL), the University of Melbourne De Brettville Trust (RW) and the Kel and Rosie Day Foundation (RW). JCS is supported by GOSHCC. The Centre for Eye Research Australia receives operational infrastructure support from the Victorian Government.

Competing Interests statement

The authors have no conflict of interest to declare.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Hornan, D. M. *et al.* Novel retinal and cone photoreceptor transcripts revealed by human macular expression profiling. *Invest. Ophthalmol. Vis. Sci.* **48,** 5388–5396 (2007).
2. Mustafi, D. *et al.* Transcriptome analysis reveals rod/cone photoreceptor specific signatures across mammalian retinas. *Hum. Mol. Genet.* **25,** 4376–4388 (2016).
3. Whitmore, S. S. *et al.* Transcriptomic analysis across nasal, temporal, and macular regions of human neural retina and RPE/choroid by RNA-Seq. *Exp. Eye Res.* **129,** 93–106 (2014).
4. Farkas, M. H. *et al.* Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics* **14,** 486 (2013).
5. Pinelli, M. *et al.* An atlas of gene expression and gene co-regulation in the human retina. *Nucleic Acids Res.* **44,** 5773–5784 (2016).
6. Hornan, D. M. *et al.* Novel retinal and cone photoreceptor transcripts revealed by human macular expression profiling. *Invest. Ophthalmol. Vis. Sci.* **48,** 5388–5396 (2007).
7. Mustafi, D. *et al.* Transcriptome analysis reveals rod/cone photoreceptor specific signatures across mammalian retinas. *Hum. Mol. Genet.* **25,** 4376–4388 (2016).
8. Whitmore, S. S. *et al.* Transcriptomic analysis across nasal, temporal, and macular regions of human neural retina and RPE/choroid by RNA-Seq. *Exp. Eye Res.* **129,** 93–106 (2014).
9. Farkas, M. H. *et al.* Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics* **14,** 486 (2013).
10. Pinelli, M. *et al.* An atlas of gene expression and gene co-regulation in the human retina. *Nucleic Acids Res.* **44,** 5773–5784 (2016).
11. Hoshino, A. *et al.* Molecular Anatomy of the Developing Human Retina. *Dev. Cell* **43,** 763–779.e4 (2017).
12. Kozulin, P., Natoli, R., O'Brien, K. M. B., Madigan, M. C. & Provis, J. M. Differential expression of anti-angiogenic factors and guidance genes in the developing macula. *Mol. Vis.* **15,** 45–59 (2009).
13. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
14. Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166,** 1308–1323.e30 (2016).
15. Phillips, M. J. *et al.* A Novel Approach to Single Cell RNA-Sequence Analysis Facilitates In Silico Gene Reporting of Human Pluripotent Stem Cell-Derived Retinal Cell Types. *Stem Cells* **36,** 313–324 (2018).
16. Welby, E. *et al.* Isolation and Comparative Transcriptome Analysis of Human Fetal and iPSC-Derived Cone Photoreceptor Cells. *Stem Cell Reports* **9,** 1898–1915 (2017).
17. Strunnikova, N. V. *et al.* Transcriptome analysis and molecular signature of human retinal pigment epithelium. *Hum. Mol. Genet.* **19,** 2468–2486 (2010).
18. Liao, J.-L. *et al.* Molecular signature of primary retinal pigment epithelium and stem-cell-derived RPE cells. *Hum. Mol. Genet.* **19,** 4229–4238 (2010).
19. Klimova, L., Antosova, B., Kuzelova, A., Strnad, H. & Kozmik, Z. Onecut1 and Onecut2 transcription factors operate downstream of Pax6 to regulate horizontal cell development. *Dev. Biol.* **402,** 48–60 (2015).
20. Vecino, E., Rodriguez, F. D., Ruzafa, N., Pereiro, X. & Sharma, S. C. Glia-neuron

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

interactions in the mammalian retina. *Prog. Retin. Eye Res.* **51,** 1–40 (2016).

21. Soto, I. *et al.* Retinal Ganglion Cells Downregulate Gene Expression and Lose Their Axons within the Optic Nerve Head in a Mouse Glaucoma Model. *Journal of Neuroscience* **28,** 548–561 (2008).
22. Imanishi, Y. *et al.* Characterization of retinal guanylate cyclase-activating protein 3 (GCAP3) from zebrafish to man. *Eur. J. Neurosci.* **15,** 63–78 (2002).
23. Blackshaw, S., Fraioli, R. E., Furukawa, T. & Cepko, C. L. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. *Cell* **107,** 579–589 (2001).
24. Corbo, J. C., Myers, C. A., Lawrence, K. A., Jadhav, A. P. & Cepko, C. L. A typology of photoreceptor gene expression patterns in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 12069–12074 (2007).
25. Jeon, C. J., Strettoi, E. & Masland, R. H. The major cell populations of the mouse retina. *J. Neurosci.* **18,** 8936–8946 (1998).
26. Boon, C. J. F. *et al.* The spectrum of retinal dystrophies caused by mutations in the peripherin/RDS gene. *Prog. Retin. Eye Res.* **27,** 213–235 (2008).
27. Wan, P., Su, W. & Zhuo, Y. Precise long non-coding RNA modulation in visual maintenance and impairment. *J. Med. Genet.* **54,** 450–459 (2017).
28. Louie, C. M. *et al.* AHI1 is required for photoreceptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. *Nat. Genet.* **42,** 175–180 (2010).
29. Takata, K.-I. Drosophila Damaged DNA-Binding Protein 1 Is an Essential Factor for Development. *Genetics* **168,** 855–865 (2004).
30. Fuller-Pace, F. V. The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim. Biophys. Acta* **1829,** 756–763 (2013).
31. Parmeggiani, F. *et al.* Retinitis Pigmentosa: Genes and Disease Mechanisms. *Curr. Genomics* **12,** 238–249 (2011).
32. Oh, E. C. T. *et al.* Rod differentiation factor NRL activates the expression of nuclear receptor NR2E3 to suppress the development of cone photoreceptors. *Brain Res.* **1236,** 16–29 (2008).
33. Roorda, A. & Williams, D. R. The arrangement of the three cone classes in the living human eye. *Nature* **397,** 520–522 (1999).
34. Fernández-Sánchez, L., Lax, P., Campello, L., Pinilla, I. & Cuenca, N. Astrocytes and Müller Cell Alterations During Retinal Degeneration in a Transgenic Rat Model of Retinitis Pigmentosa. *Front. Cell. Neurosci.* **9,** 484 (2015).
35. Subirada, P. V. *et al.* A journey into the retina: Müller glia commanding survival and death. *Eur. J. Neurosci.* **47,** 1429–1443 (2018).
36. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25,** 1860–1872 (2015).
37. Lawrence, J. M. *et al.* MIO-M1 cells and similar muller glial cell lines derived from adult human retina exhibit neural stem cell characteristics. *Stem Cells* **25,** 2033–2043 (2007).
38. Wirsching, H.-G. *et al.* Thymosin beta 4 gene silencing decreases stemness and invasiveness in glioblastoma. *Brain* **137,** 433–448 (2013).
39. Yoshida, S. *et al.* Expression profiling of the developing and mature Nrl −/− mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl. *Hum. Mol. Genet.* **13,** 1487–1503 (2004).

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

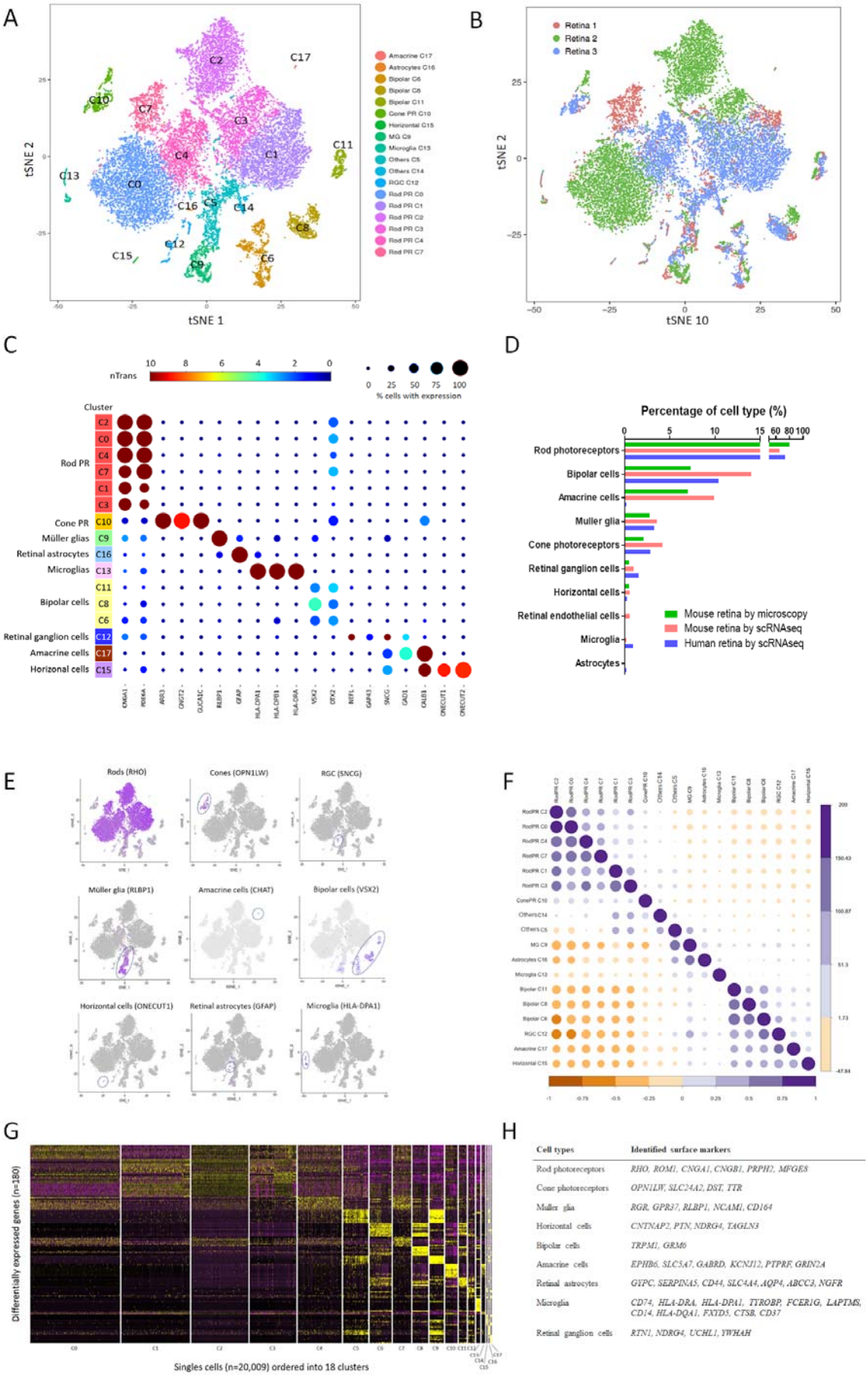
40. Gill, K. P. *et al.* Enriched retinal ganglion cells derived from human embryonic stem cells. *Sci. Rep.* **6,** 30552 (2016).
41. Kobayashi, W. *et al.* Culture Systems of Dissociated Mouse and Human Pluripotent Stem Cell-Derived Retinal Ganglion Cells Purified by Two-Step Immunopanning. *Invest. Ophthalmol. Vis. Sci.* **59,** 776–787 (2018).
42. Sluch, V. M. *et al.* Differentiation of human ESCs to retinal ganglion cells using a CRISPR engineered reporter cell line. *Sci. Rep.* **5,** 16595 (2015).
43. Lakowski, J. *et al.* Isolation of Human Photoreceptor Precursors via a Cell Surface Marker Panel from Stem Cell-Derived Retinal Organoids and Fetal Retinae. *Stem Cells* (2018). doi:10.1002/stem.2775
44. Li, H.-B. *et al.* Long Non-Coding RNA-MALAT1 Mediates Retinal Ganglion Cell Apoptosis Through the PI3K/Akt Signaling Pathway in Rats with Glaucoma. *Cell. Physiol. Biochem.* **43,** 2117–2132 (2017).
45. Yang, S., Yao, H., Li, M., Li, H. & Wang, F. Long Non-Coding RNA MALAT1 Mediates Transforming Growth Factor Beta1-Induced Epithelial-Mesenchymal Transition of Retinal Pigment Epithelial Cells. *PLoS One* **11,** e0152687 (2016).
46. Tsukamoto, Y., Morigiwa, K., Ueda, M. & Sterling, P. Microcircuits for night vision in mouse retina. *J. Neurosci.* **21,** 8616–8623 (2001).
47. Kim, J.-W. *et al.* Recruitment of Rod Photoreceptors from Short-Wavelength-Sensitive Cones during the Evolution of Nocturnal Vision in Mammals. *Dev. Cell* **37,** 520–532 (2016).
48. Viets, K., Eldred, K. & Johnston, R. J., Jr. Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates. *Trends Genet.* **32,** 638–659 (2016).
49. Hung, S. S. C., Khan, S., Lo, C. Y., Hewitt, A. W. & Wong, R. C. B. Drug discovery using induced pluripotent stem cell models of neurodegenerative and ocular diseases. *Pharmacol. Ther.* **177,** 32–43 (2017).
50. Handel, A. E. *et al.* Assessing similarity to primary tissue and cortical layer identity in induced pluripotent stem cell-derived cortical neurons through single-cell transcriptomics. *Hum. Mol. Genet.* **25,** 989–1000 (2016).
51. Baxter, M. *et al.* Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. *J. Hepatol.* **62,** 581–589 (2015).
52. Hollborn, M. *et al.* The human Müller cell line MIO-M1 expresses opsins. *Mol. Vis.* **17,** 2738–2750 (2011).
53. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** 14049 (2017).
54. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
55. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17,** 75 (2016).
56. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).
57. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44,** W90–7 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure Legends

Figure 1: Single cell transcriptome atlas for human neural retina. t-SNE visualization of 20,009 human retinal cells colored by A) annotation of 18 transcriptionally distinct clusters (C0-C17) and B) their distribution in 3 donor retina samples. C) Feature expression heatmap showing expression patterns of major retinal class markers across 16 retinal cell clusters. The size of each circle depicts the percentage of cells expressing the marker within the cluster. Brown colour indicates ≥10 nTrans (number of transcripts). D) Fraction of major cell types in human retina were estimated by the cell number in each annotated clusters and compared to previous data in mouse retina obtained by microscopy techniques [25] and scRNA-seq [13]. E) t-SNE plots showing expression of a set of selected marker genes for major retinal classes. F) Correlation matrix for the identified 18 clusters. The upper triangle depicts the z-value for correlation and the lower triangle depicts the correlation coefficient for gene expression in clusters. G) Heatmap of differentially expressed genes used to classify cell types for each cluster compared to all other clusters for the 18 retinal cell clusters. The rows correspond to top 10 genes most selectively upregulated in individual clusters ($p < 0.01$, Benjamini-Hochberg correction) and the columns show individual cells ordered by cluster (C0-C17). H) Membrane-related markers were extracted from GO annotation using the differentially expressed genes in the major human retinal cell types.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

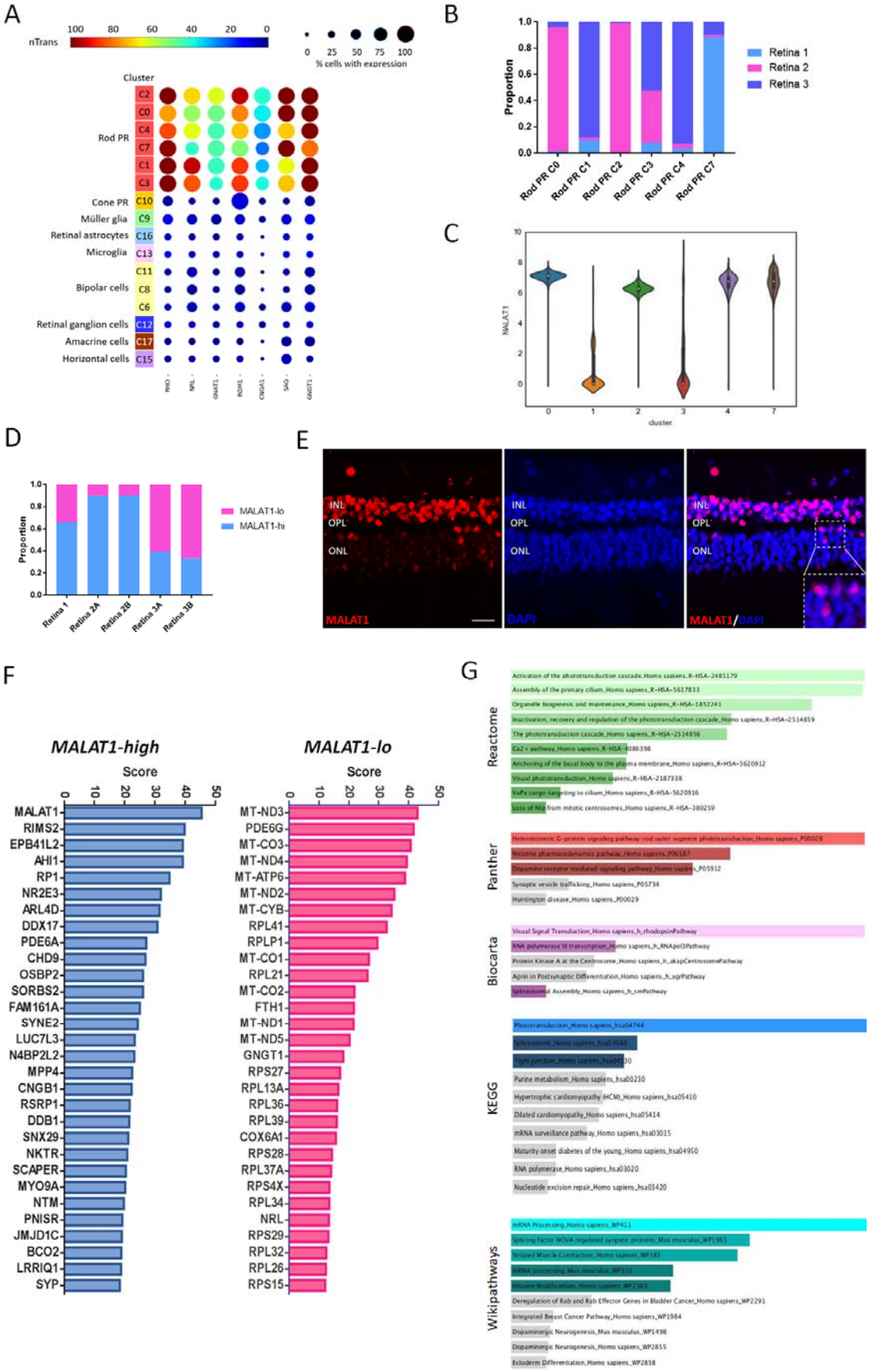


Cell types	Identified surface markers
Rod photoreceptors	*RHO, ROM1, CNGA1, CNGB1, PRPH2, MFGE8*
Cone photoreceptors	*OPN1LW, SLC24A2, DST, TTR*
Muller glia	*RGR, GPR37, RLBP1, NCAM1, CD164*
Horizontal cells	*CNTNAP2, PTN, NDRG4, TAGLN3*
Bipolar cells	*TRPM1, GRM6*
Amacrine cells	*EPHB6, SLC5A7, GABRD, KCNJ12, PTPRF, GRIN2A*
Retinal astrocytes	*GYPC, SERPINA5, CD44, SLC4A4, AQP4, ABCC3, NGFR*
Microglia	*CD74, HLA-DRA, HLA-DPA1, TYROBP, FCER1G, LAPTM5, CD14, HLA-DQA1, FXYD5, CTSB, CD37*
Retinal ganglion cells	*RTN1, NDRG4, UCHL1, YWHAH*

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2: Identification of *MALAT1-hi* and *MALAT1-lo* subpopulations of rod photoreceptors. A) Feature expression heatmap of a panel of known marker genes for rod photoreceptors across the identified 16 retinal cell clusters. Brown colour indicates ≥100 nTrans (number of transcripts).B) Representation of the three donor retina samples in the six rod photoreceptor clusters. C) Violin plot showing high or low expression levels of *MALAT1* in rod photoreceptor clusters. D) Distribution of rod photoreceptor populations with high MALAT1 expression (*MALAT1*-hi) or low *MALAT1* expression (*MALAT1*-lo) in three donor retina samples. E) Fluorescent in situ hybridization analysis of human peripheral retina showing heterogeneous levels of *MALAT1* expression in the rod photoreceptors located in the outer nuclear layer (ONL). INL: inner nuclear layer; OPL: outer plexiform layer. Scale bar = 20µm. F) Top 30 differentially expressing genes in rod photoreceptors with high (*MALAT1-hi*) or low *MALAT1* expression levels (*MALAT1-lo*) ranked by logistic regression score. G) Gene enrichment analysis of 100 genes differentially up-regulated in high *MALAT1* expressing cells (*MALAT1-hi*) compared to low *MALAT1* expressing cells (*MALAT1-lo*) for metabolic and signaling pathway analysis (Reactome, Panther, Biocarta, KEGG, Wikipathways). The bars depict the combined score of the log of p-value from the Fisher exact test and the z-score of the deviation of the expected rank [57].

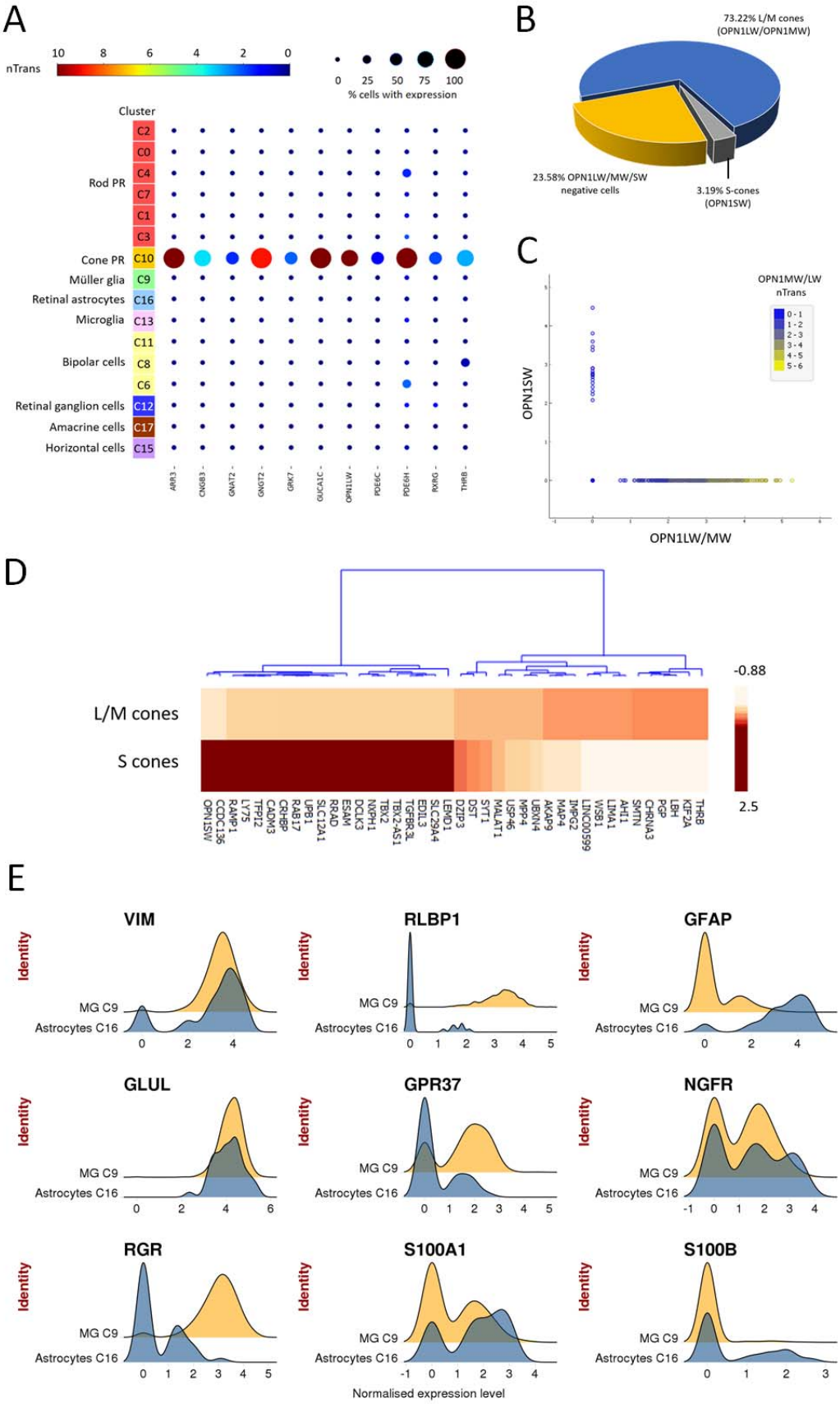
bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3: Assessment of cone photoreceptor and glial cell types in human retina . A) Feature expression heatmap showing the expression of 11 known cone photoreceptor markers across 16 retinal cell clusters. Brown colour indicates ≥10 nTrans (number of transcripts). B) The proportion of cone photoreceptor subtypes identified in C10 cluster, based on expression of *OPN1LW / OPN1MW* (L/M cones) and *OPN1SW* (S cones). C) Scatter plots showing expression of *OPN1LW/OPN1MW* or *OPN1SW* in individual cone photoreceptors for C10 cluster. The colour depicts expression level for *OPN1LW/OPN1MW* in individual cells. D) Heatmap of top 20 differentially expressed genes between L/M cones and S cones. The colour depicts normalised gene expression (z-score capped at 2.5). E) Expression pattern of glial markers in Muller glia (C9) and retinal astrocytes (C16). x-axis depicts normalized transcript levels.

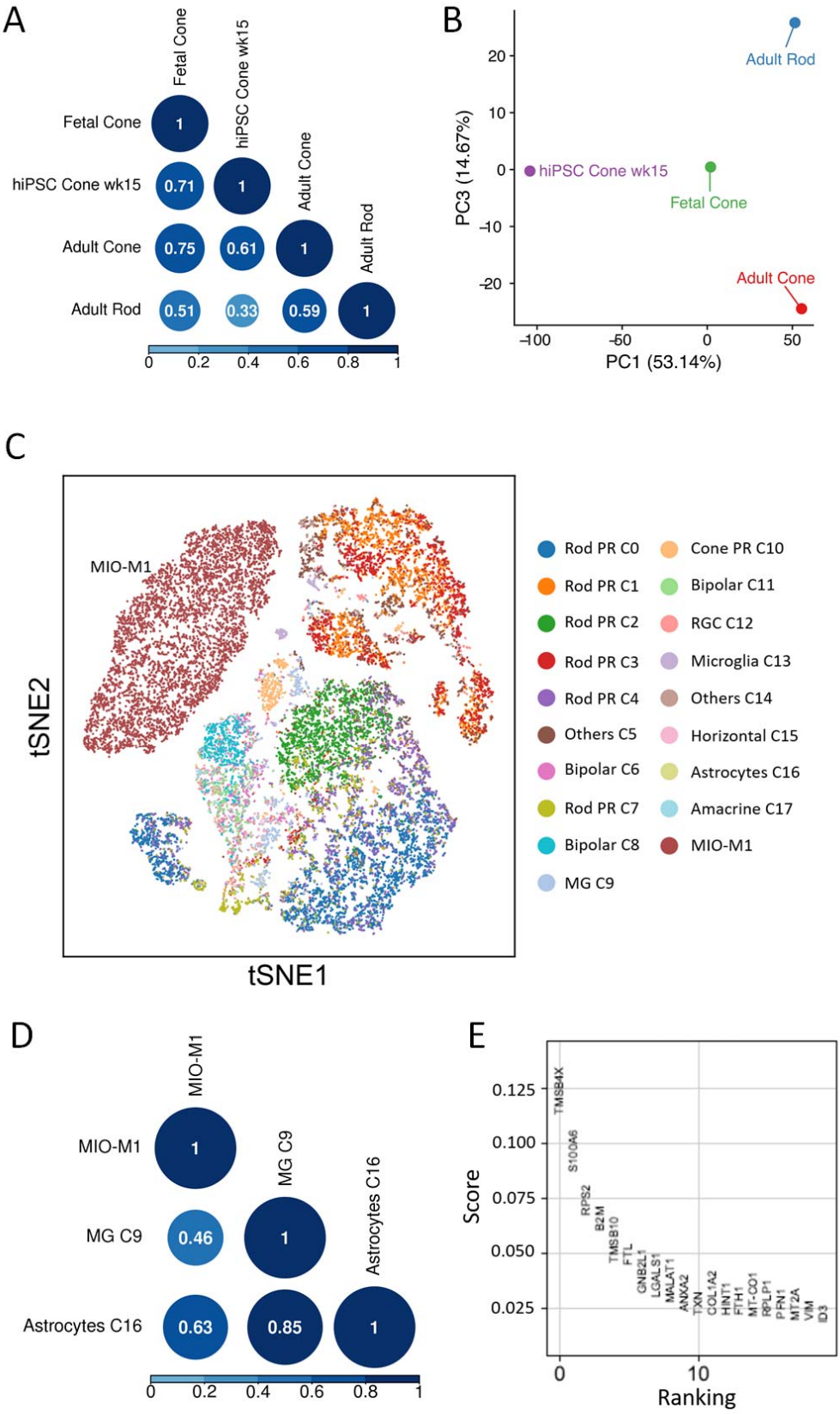
bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4: Benchmarking retinal cells using the human neural retina atlas. A) Correlation analysis of scRNA-seq data of hiPSC-derived cone photoreceptors (week 15) against fetal cone photoreceptors [16], as well as adult cone and rod photoreceptors from this human neural retina atlas. B) Principal component analysis to assess transcriptome similarity of hiPSC-derived cone photoreceptors to fetal and adult photoreceptors. C) t-SNE analysis of the human MG cell line MIO-M1 with the retinal cell types identified in this human neural retina atlas. D) Correlation analysis of MIO-M1 with all major human retinal cell types. E) Top ranked differentially expressed genes identified in MIO-M1 compared to other retinal cell types based on logistic regression score.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary figures and tables:

Supplementary table 1: Details for donor retina samples.

Retina	Patient ID	Eye bulb	Sex	Age (years)	Retrieval time (hrs)	Ocular complications	Assay	scRNA-seq Library	Targeted cell number	Captured cell number after QC
1	17-010	Right	F	80	11.5	Cataract on left eye	scRNAseq	A	4000	2122
2	SC	Left	M	42	6.2	-	scRNAseq	A	10000	4449
								B	10000	4528
3	17-011	Right	F	53	14.5	-	scRNAseq	A	10000	4518
								B	10000	4392
4	16033	Left	M	73	12	-	FISH			
5	16061	Left	M	79	25	-	FISH			
6	16088	Left	F	61	22	-	FISH			

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary table 2: Breakdown of cell number assigned to individual donor libraries. The red highlight the number of cells assigned in the largest clusters, showing similar cell assignment to clusters between 2 replicates.

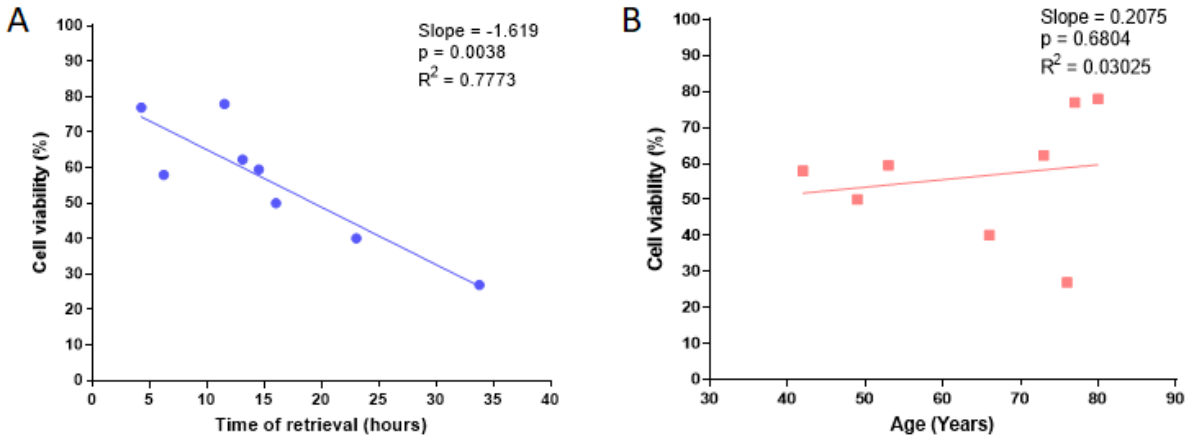
Clusters	Retina 1	Retina 2A	Retina 2B	Retina 3A	Retina 3B
0	58	1927	1924	103	66
1	307	33	30	1419	1330
2	3	1275	1299	9	9
3	172	413	425	447	670
4	74	23	46	999	864
5	232	59	74	398	384
6	156	170	139	273	264
7	737	6	8	42	41
8	97	98	124	208	190
9	106	184	189	93	86
10	33	83	80	180	188
11	42	83	68	93	77
12	55	35	35	108	82
13	18	20	40	58	51
14	14	7	11	62	71
15	11	4	12	13	9
16	5	8	11	9	9
17	2	21	13	4	1

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary table 3: List of top 10 differential gene expression in rod photoreceptor clusters

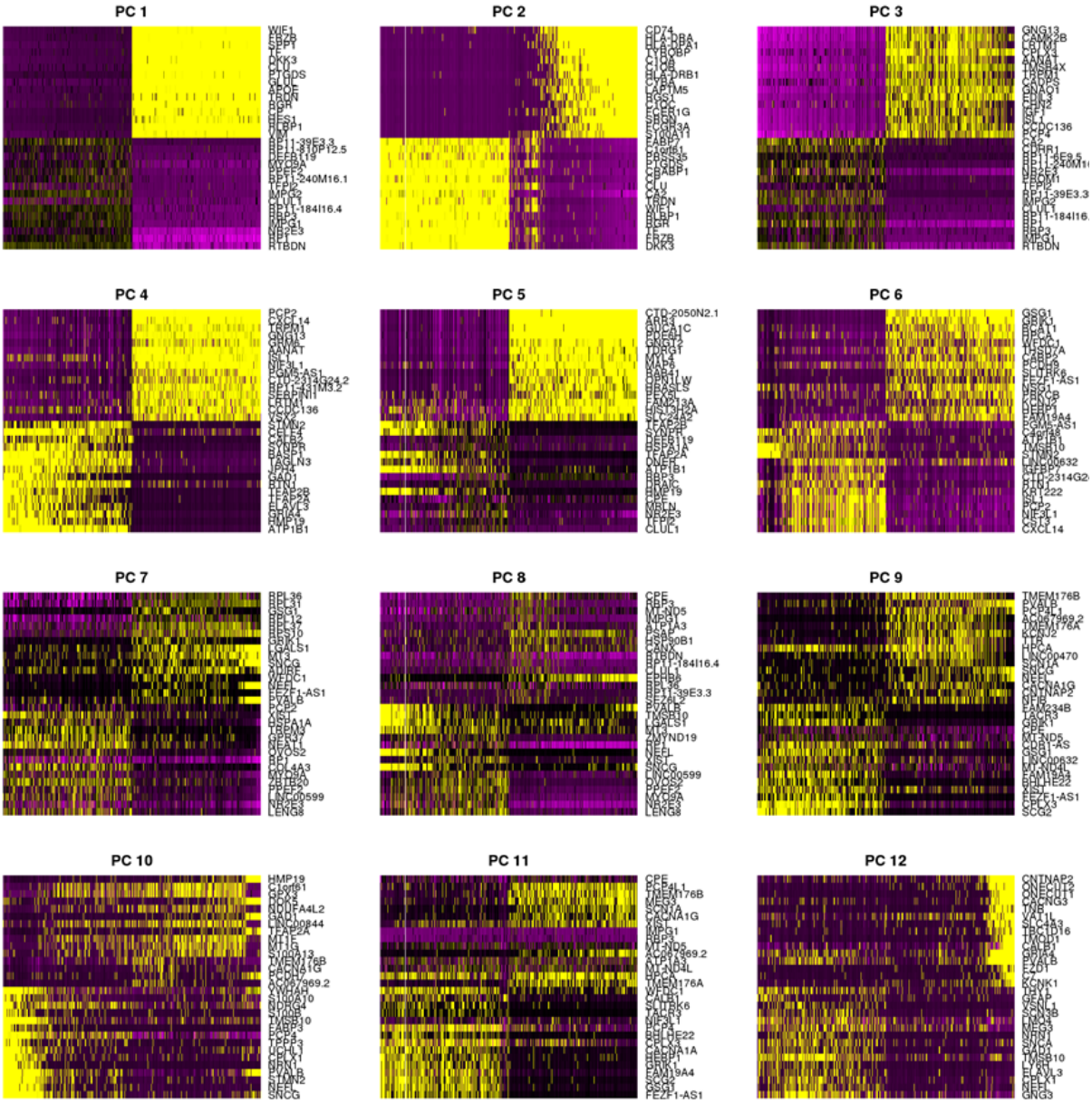
Cluster	**Top 10 differential genes**
Rod PR C0	*NR2E3, ARL4D, AHI1, OSBP2, EPB41L2, RP1, RIMS2, FAM161A, SCAPER, SORBS2*
Rod PR C1	*IMPG1, RBP3, PDE6G, COX6A1, CPE, RP11-184I16.4, RP11-6E9.5, RP11-e9E3.3, SYCE1L, ATP1A3*
Rod PR C2	*HSP90AA1, RPS4Y1, SAG, ALDOC, RBP3, ARL4D, CLUL1, CKB, KCNV2, ELOVL4*
Rod PR C3	*MT-ND3, MT-ATP6, MT-ND2, MT-CO3, MT-ND5, MT-ND4, MT-CYB, MT-ND1, MT-CO2, MT-ND4L*
Rod PR C4	*MYO9A, XIST, NTM, NR2E3, RP11-798M19.6, ZMYND19, OVOS2, LENG8, LINC00599, BCO2*
Rod PR C7	*LENG8, HSPA1B, PPEF2, HSPA1A, XIST, MPP4, DMD, OVOS2, ABCA4, JUN*

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



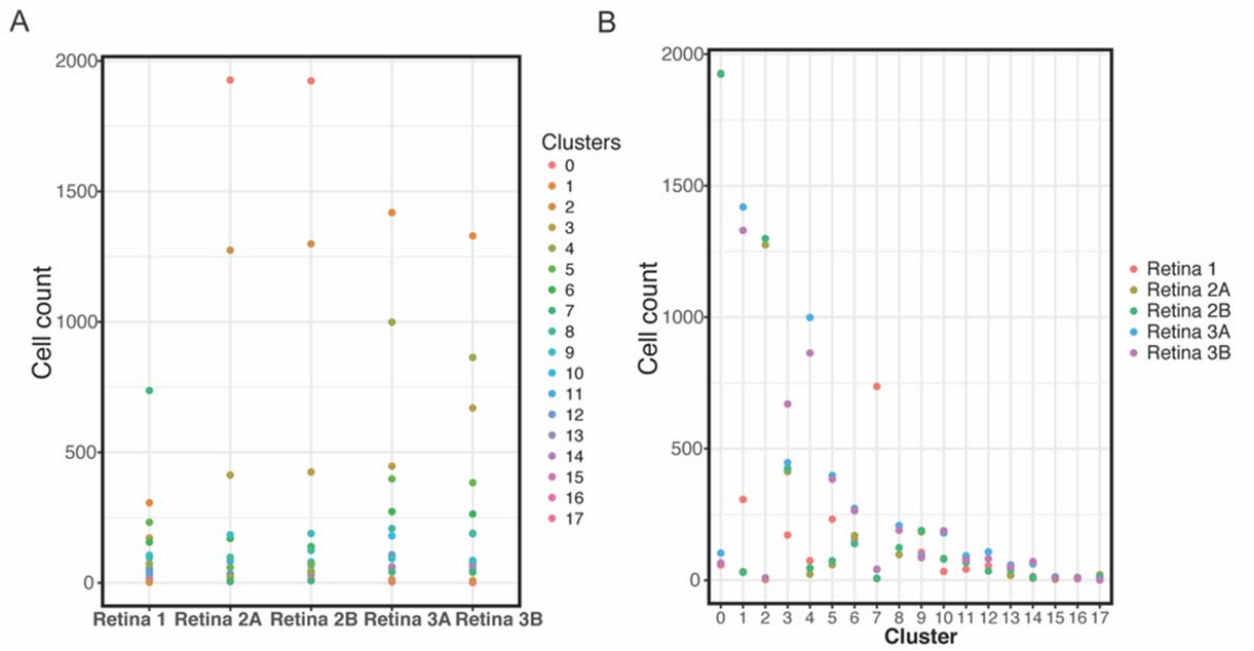
Supplementary figure 1: Correlation of cell viability of post-mortem human neural retina with A) time of tissue retrieval after death, and B) age of donor.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



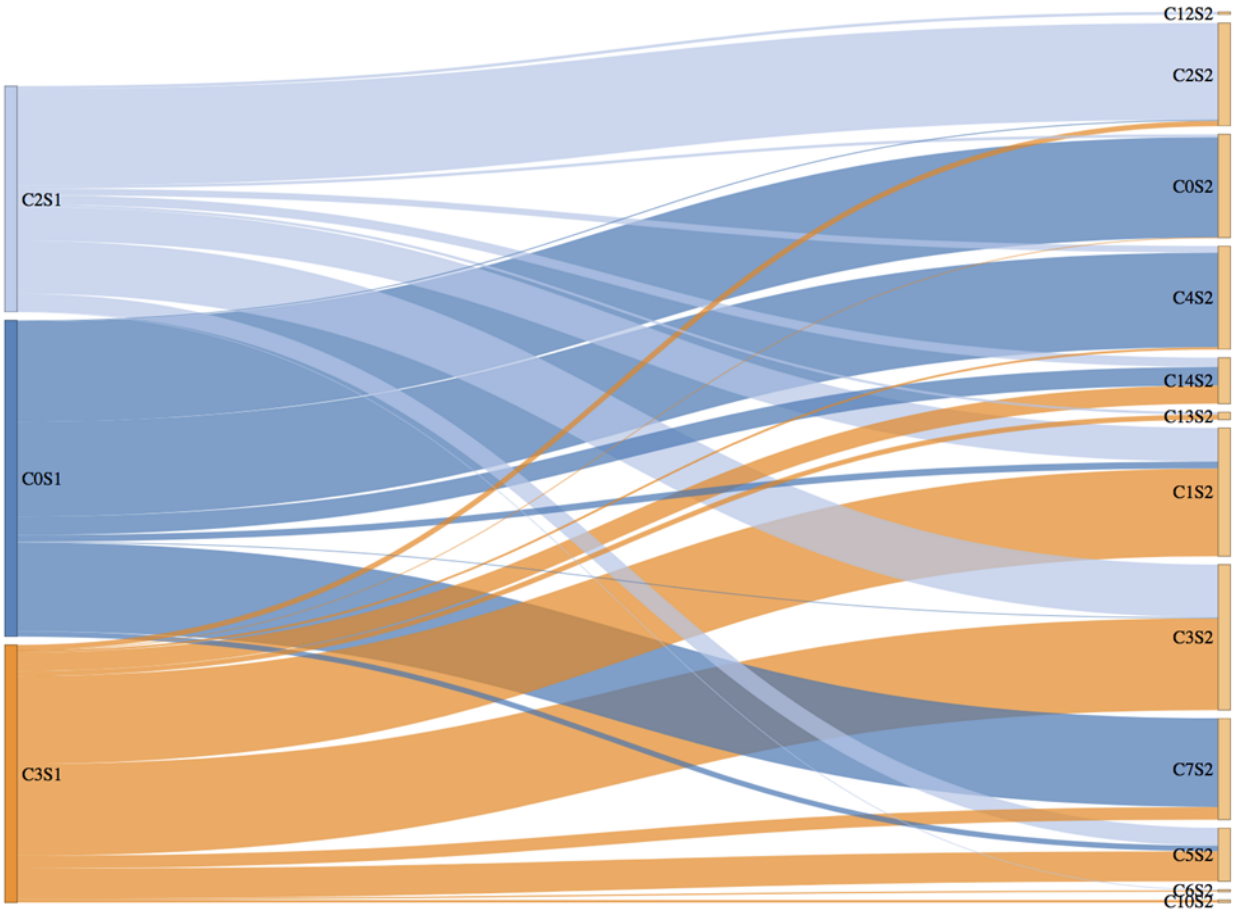
Supplementary figure 2: Heatmaps of the top 12 principal components explaining the primary sources of heterogeneity in the retinal scRNA-seq data. Cells and genes are ordered by PCA score calculated by Seurat. The genes driving the majority of the variance are determined using the top 500 cells.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



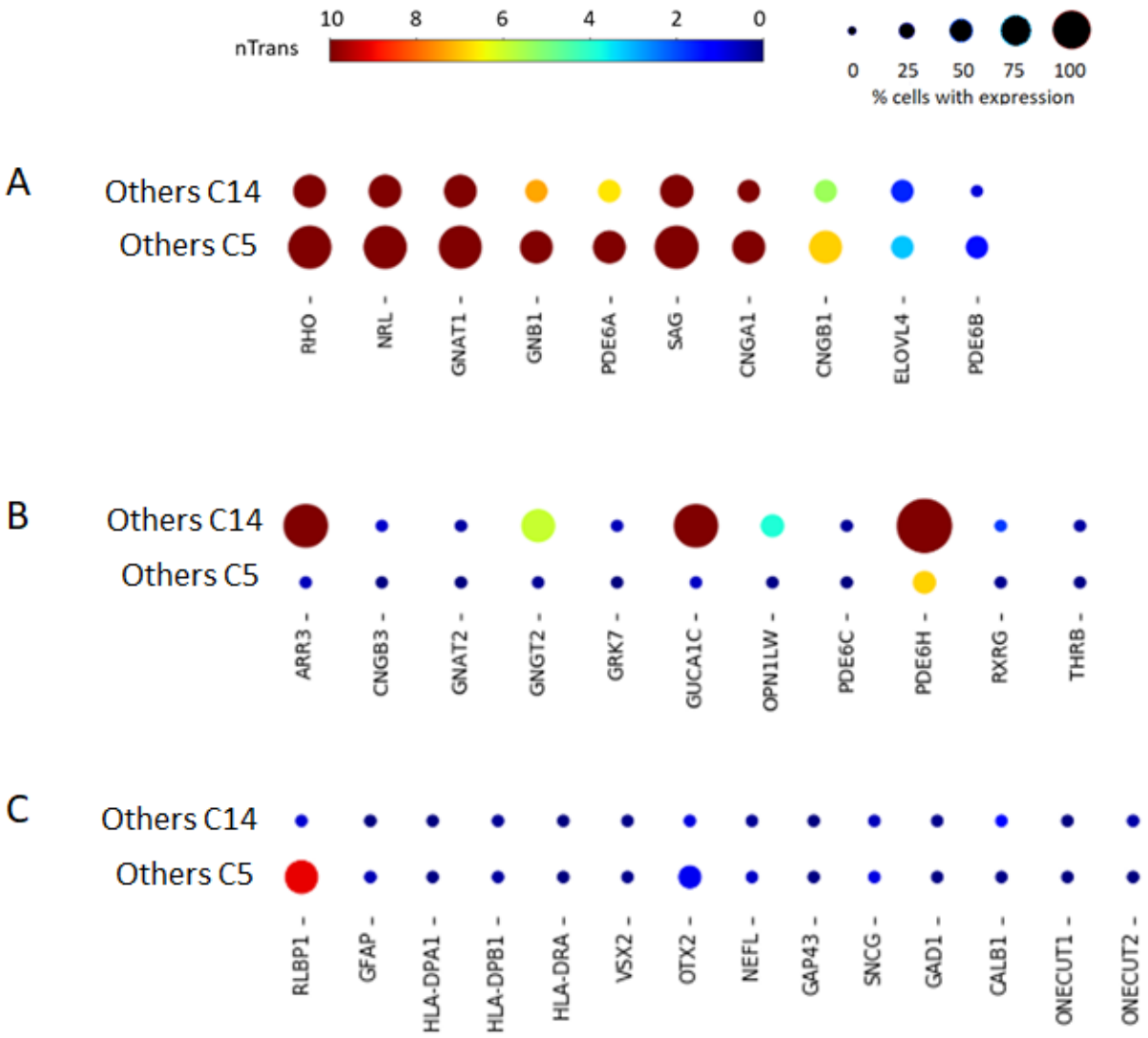
Supplementary figure 3: Distribution of frequency of the 18 clusters (C0-C17) in individual single cell libraries ordered by A) individual single cell library or B) identified clusters.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



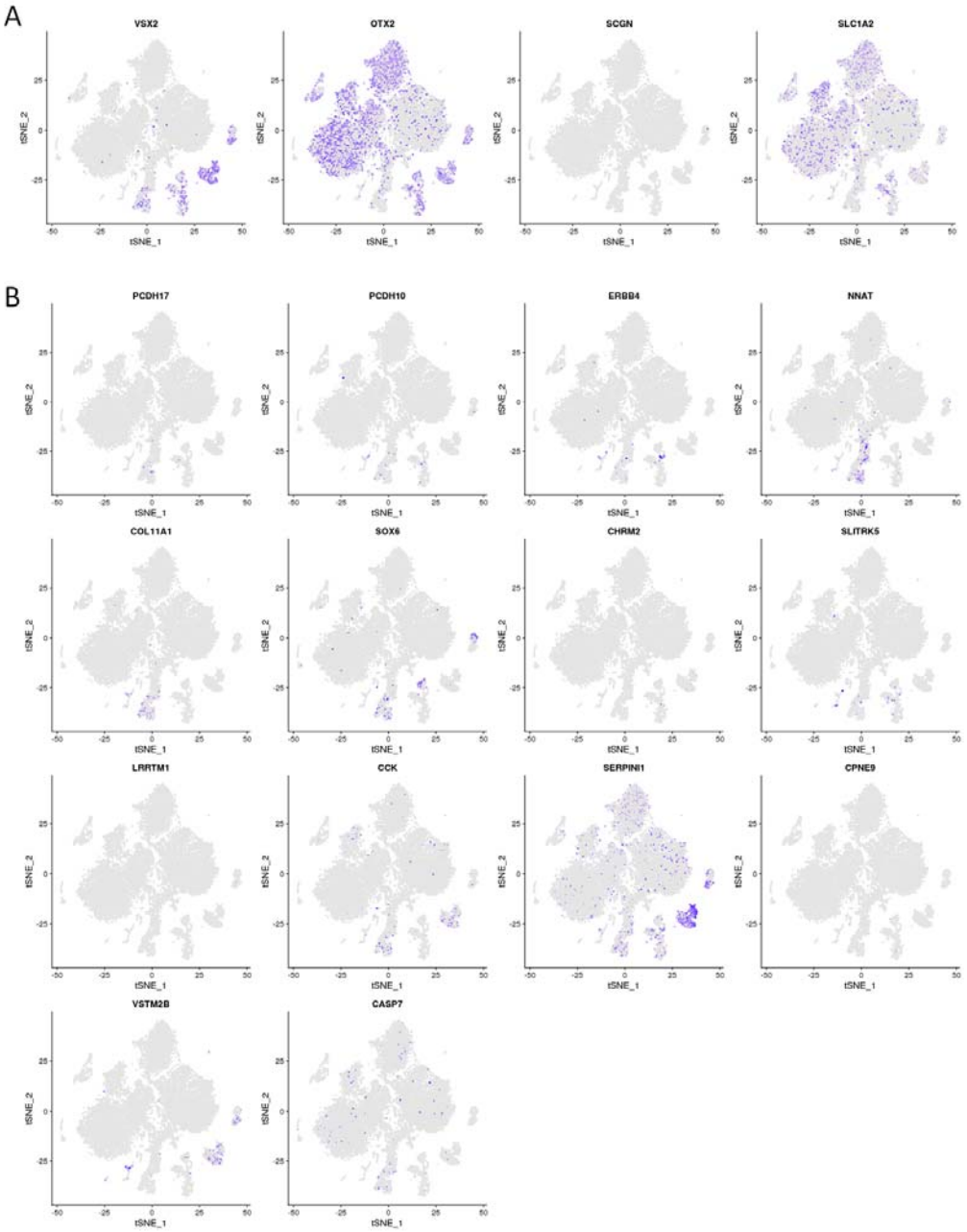
Supplementary figure 4: Prediction of cluster relationship between library technical replicates in Retina 1 and Retina 2. The Sankey plot shows edges connecting clusters, with larger edge indicating higher similarity, ranging from 0 to 100%. The size of the edge was quantitatively estimated by implementing scGPS modelling approach for pairs of clusters, as described in the method section. The three largest clusters in Retina 2A were compared with all clusters in Retina 2B. Consistently we see C2 in Retina 2A is most similar to C2 in Retina 2B. The same trend is seen for C0 and C3. These results demonstrated that the variation between library replicates is minimal in our dataset, and that the clusters determined from the merged dataset were consistent across samples. We also found higher similarities among Rod photoreceptor clusters (C0, 2, 3 in Retina 2A with clusters C0, 2, 4, 7 in Retina 2B) than compared with other clusters.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



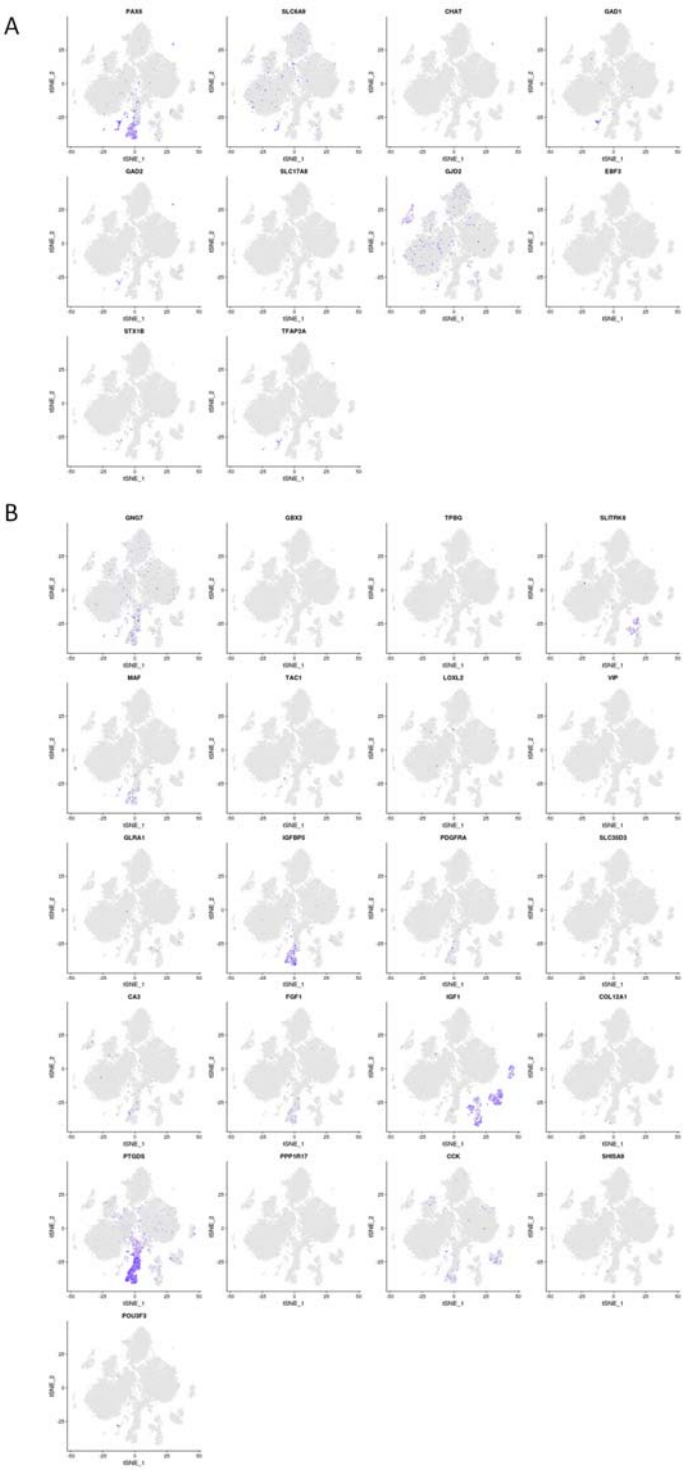
Supplementary figure 5: Feature expression heatmap showing expression patterns of A) rod photoreceptor markers, B) cone photoreceptor markers and C) other major retinal class markers (MG: RLBP1; astrocytes: GFAP; microglia: HLA-DPA1, HLA-DPB1, HLA-DRA; Bipolar cells: VSX2, OTX2; retinal ganglion cells: NEFL, GAP43, SNCG; Amacrine cells: GAD1, CALB1; Horizontal cells: ONECUT1, ONECUT2) in unassigned clusters C14 and C5. The size of each circle depicts the percentage of cells expressing the marker within the cluster.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



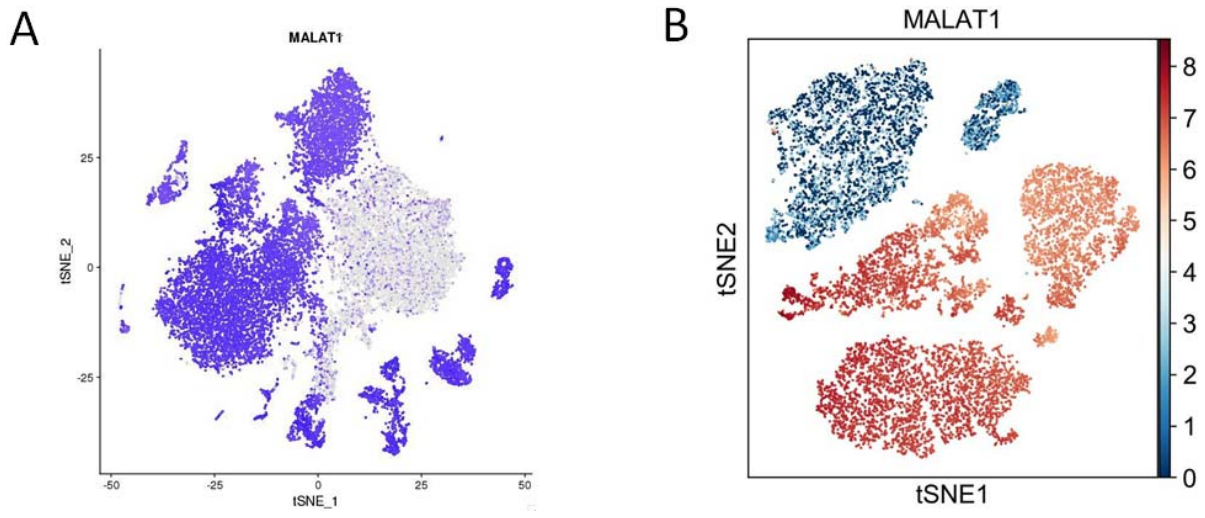
Supplementary figure 6: t-SNE plots showing gene expression in the compiled human neural retina transcriptome atlas (20,009 cells) for A) 4 commonly used bipolar markers and B) 14 new markers for individual bipolar subtypes identified in previous mouse scRNA-seq study [14].

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



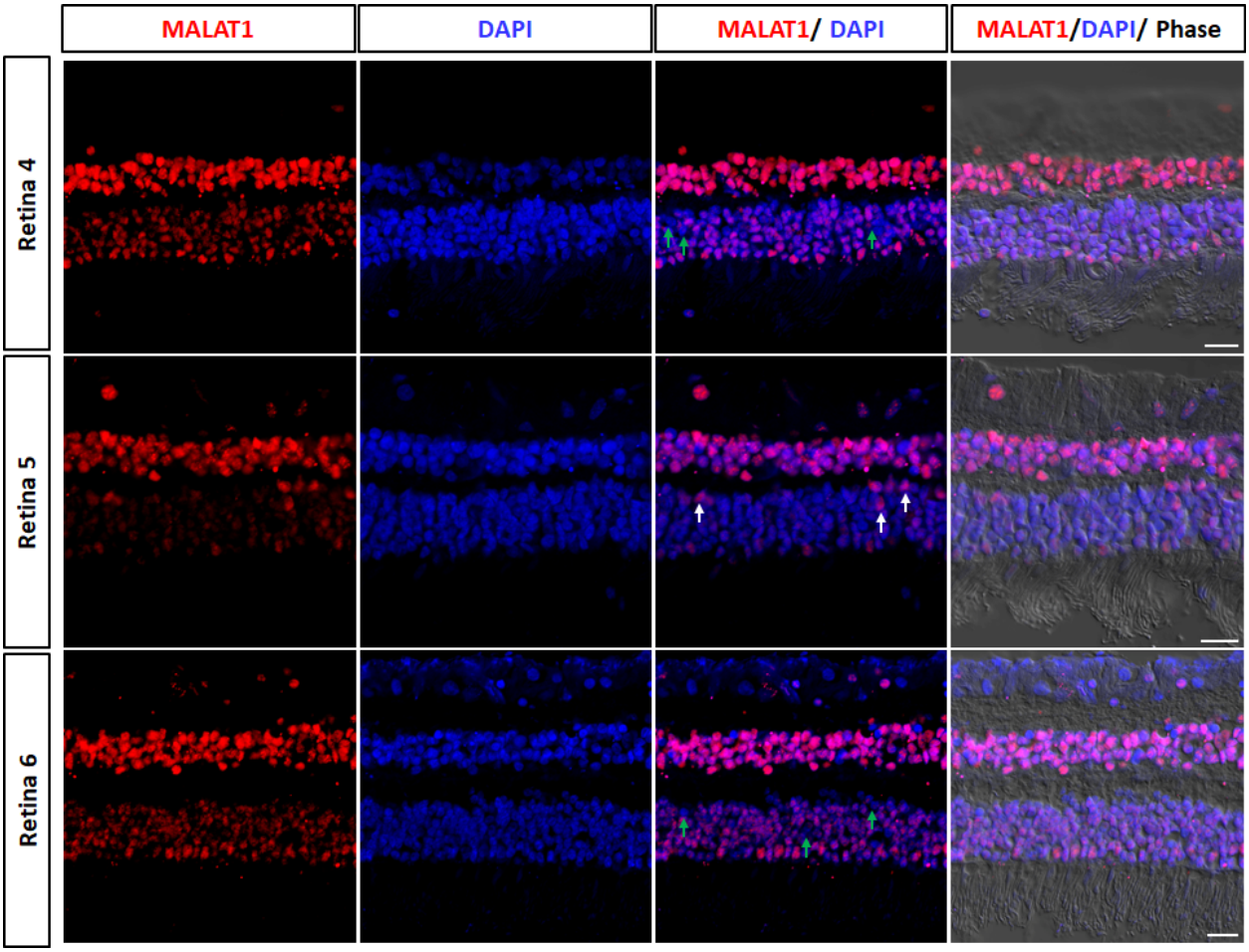
Supplementary figure 7: t-SNE plots showing gene expression in the compiled human retina transcriptome atlas (20,009 cells) for A) 10 commonly used amacrine markers and B) new markers for amacrine subtypes identified in previous mouse scRNA-seq study [13].

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



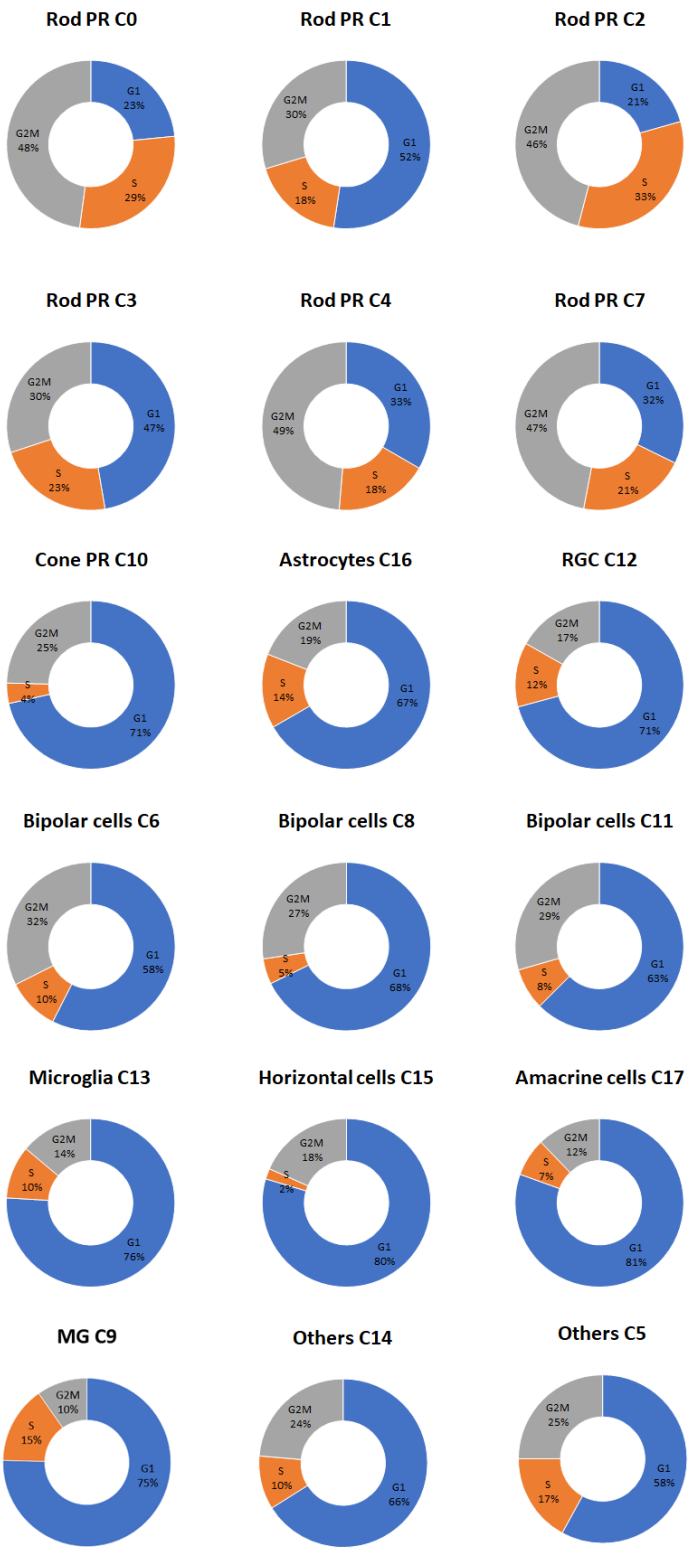
Supplementary figure 8: A) t-SNE plot showing *MALAT1* expression across all retinal cells. Blue colour indicate normalised gene expression level. B) The six rod photoreceptor clusters were pooled together for t-SNE clustering analysis and showed that the clusters can be categorized by high levels (red/orange) or low levels (blue) of *MALAT1* expression.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



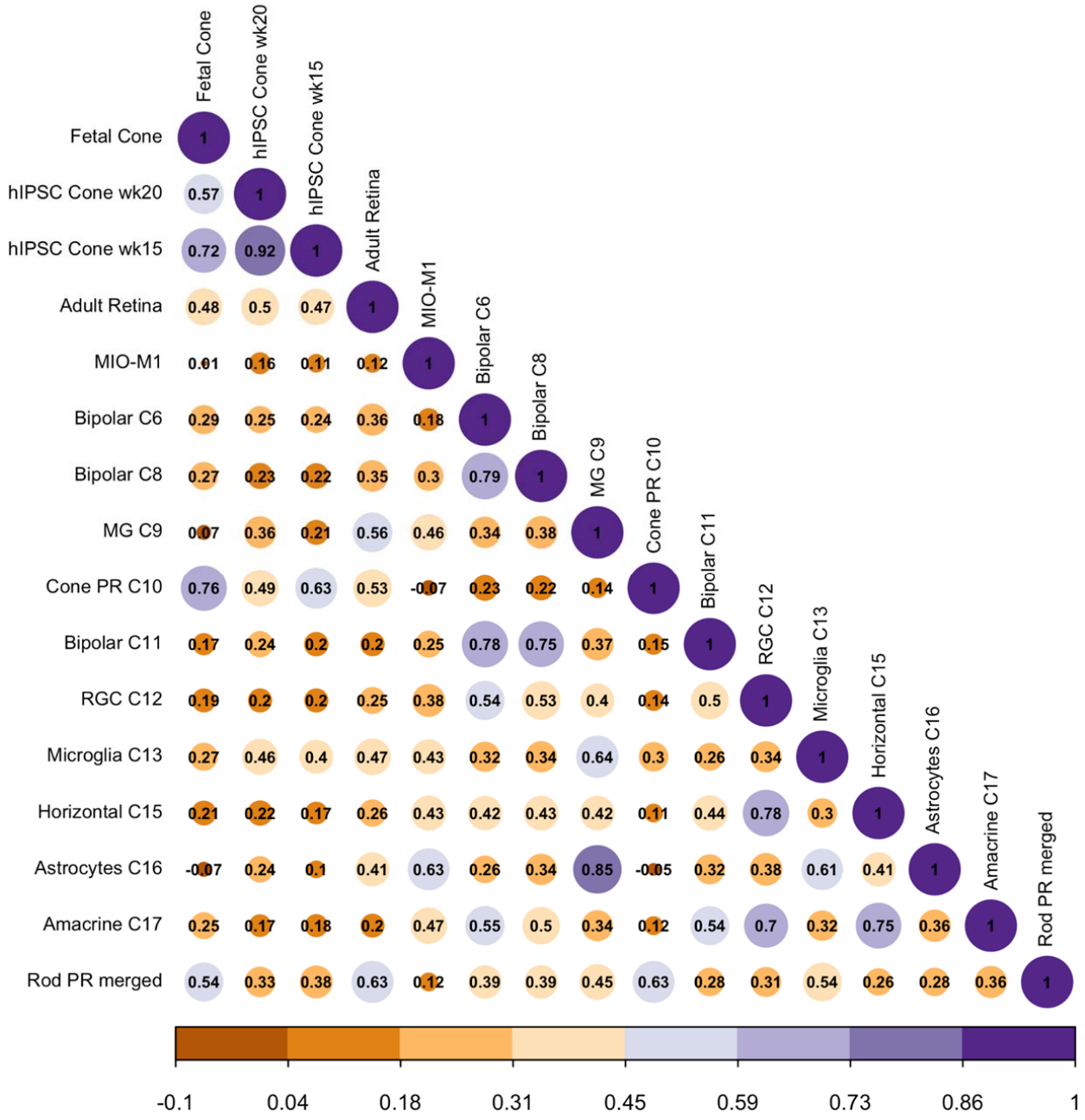
Supplementary figure 9: Fluorescent in situ hybridization showing expression of *MALAT1* in three donor retina samples (Retina 4-6). Green arrows highlight rod photoreceptors with low levels of *MALAT1* in Retina 4 and 6, white arrows highlight rod photoreceptors with high levels of *MALAT1* in Retina 5. Scale bars = 20µm

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Supplementary figure 10: Cell cycle scores across major retinal cell clusters showing the likelihood for the proportion of cells in G1, S or G2/M phases.

bioRxiv preprint doi: https://doi.org/10.1101/425223; this version posted September 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



Supplementary figure 11: Correlation matrix to benchmark hiPSC-derived cone photoreceptors (week 15, week 20; [16], fetal cone photoreceptors [16], adult retina [15] and the human Müller glia cell line MIO-M1 against all retinal cell types identified in this human neural retina atlas.