

UC Berkeley

UC Berkeley Previously Published Works

Title

Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina

Permalink

<https://escholarship.org/uc/item/92h2n8n6>

Journal

Cell, 176(5)

ISSN

0092-8674

Authors

Peng, Yi-Rong
Shekhar, Karthik
Yan, Wenjun
[et al.](#)

Publication Date

2019-02-01

DOI

10.1016/j.cell.2019.01.004

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed



Published in final edited form as:

Cell. 2019 February 21; 176(5): 1222–1237.e22. doi:10.1016/j.cell.2019.01.004.

MOLECULAR CLASSIFICATION AND COMPARATIVE TAXONOMICS OF FOVEAL AND PERIPHERAL CELLS IN PRIMATE RETINA

Yi-Rong Peng^{#1}, Karthik Shekhar^{#2}, Wenjun Yan¹, Dustin Herrmann¹, Anna Sappington², Greg S. Bryman³, Tavé van Zyl⁴, Michael Tri. H. Do³, Aviv Regev^{2,5}, and Joshua R. Sanes^{1,^}

¹Center for Brain Science and Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

²Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142

³F.M. Kirby Neurobiology Center, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115

⁴Department of Ophthalmology, Harvard Medical School, Boston, MA 02114

⁵Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02140

These authors contributed equally to this work.

SUMMARY

High acuity vision in primates, including humans, is mediated by a small central retinal region called the fovea. As more accessible organisms lack a fovea, its specialized function and its dysfunction in ocular diseases remain poorly understood. We used 165,000 single-cell RNA-seq profiles to generate comprehensive cellular taxonomies of macaque fovea and peripheral retina. More than 80% of >60 cell types match between the two regions, but exhibit substantial differences in proportions and gene expression, some of which we relate to functional differences. Comparison of macaque retinal types with those of mice reveals that interneuron types are tightly conserved. In contrast, projection neuron types and programs diverge, despite exhibiting conserved transcription factor codes. Key macaque types are conserved in humans, allowing mapping of cell-type and region-specific expression of >190 genes associated with 7 human retinal diseases. Our work provides a framework for comparative single-cell analysis across tissue regions and species.

[^]Lead contact: Joshua R. Sanes (sanesj@mcb.harvard.edu)

AUTHOR CONTRIBUTIONS

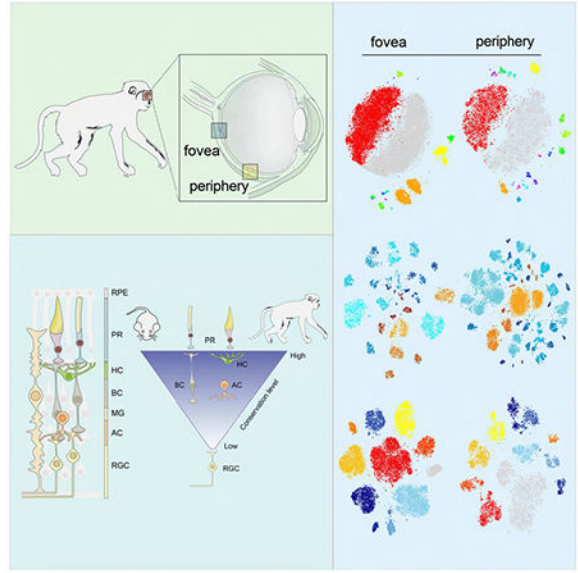
Y.-R.P., K.S., A. R., and J.R.S conceived the study and wrote the manuscript with input from all authors. K.S., W.Y. and A.S. performed bioinformatic analysis, with guidance from A. R. Y-R.P. and D.H. performed molecular and histological experiments, with guidance from M. T. H. D and J. R. S.. T.v.T. and G.S.B. procured and prepared tissue. M.T.H.D. provided guidance on study design and analysis.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

COMPETING INTERESTS

AR is an equity holder in Celsius Therapeutics and an SAB member in Syros Pharmaceuticals and Thermo Fisher Scientific.

Graphic Abstract



INTRODUCTION

Most primates, including humans, see objects clearly only when they look straight at them so their image falls on a small central region of the retina called the fovea (Figure 1A). This is because the fovea, which may be the only primate-specific structure in the mammalian central nervous system, mediates high acuity vision. Although it occupies <1% of the retinal surface, it provides ~50% of the input to primary visual cortex (Bringmann et al., 2018; Provis et al., 2013). Diseases that selectively affect this region, such as macular degeneration, are leading causes of human blindness in the developed world (Bourne et al., 2013). These important aspects of human retinal function and dysfunction cannot be satisfactorily studied in more accessible model organisms, including mice, because they lack a fovea.

Basic aspects of retinal structure and function are shared by all retinal regions in all vertebrate species (Figure 1B). Photoreceptors (PR, rods and cones) convert light into electrical signals, which are transferred to interneurons (horizontal, bipolar and amacrine cells; HCs, BCs and ACs). Interneurons process the information and deliver it to retinal ganglion cells (RGCs), which send axons through the optic nerve to the rest of the brain (Masland, 2012). Selective connectivity among multiple AC, BC and RGC types, along with type-specific intrinsic differences, make each RGC type responsive to specific visual features (Gollisch and Meister, 2010). Despite this conserved plan, however, the fovea and peripheral retina differ in many ways (Bringmann et al., 2018). For example, many RGCs in the fovea are supplied by a single PR, which maximizes acuity, whereas peripheral RGCs are supplied by dozens to hundreds of PRs, enhancing sensitivity (Figure S1A). In addition, the fovea is rich in cones, which mediate color vision in well-lit conditions, whereas peripheral retina is dominated by rods, which function in dim light, further enhancing sensitivity (night vision). Cell types in the two regions also exhibit distinct morphologies

and physiological functions (Figure S1A) (Greeff, 1874; Kolb and Marshak, 2003; Sinha et al., 2017). Because molecular information is lacking for most primate retinal cell types, and genetic access is limited, the underpinnings of these differences are little explored.

Lack of molecular knowledge has also precluded assessment of the conservation of retinal cell types between primates and more widely-used model organisms. Identification of orthologous cell types is challenging in the absence of such knowledge. The 5 retinal neuronal classes (PR, HC, BC, AC, RGC) are divided into >120 types in mice and ~60 in primates, but the extent of their correspondence is unclear (e.g., Sanes and Masland, 2015; Masri et al., 2016). Molecular characterization could provide a lingua franca to enable mapping, informing relationships between common model organisms and primates.

To address these issues, we generated an atlas of >165,000 single-cell RNA-seq (scRNA-seq) profiles from the fovea and peripheral retina of adult crab-eating macaques (*Macaca fascicularis*), a widely used primate model in vision studies (Figure S1B). We identified and molecularly characterized >60 cell types in each region, and validated many by them by pairing molecular markers with cell morphology. We then compared the fovea and periphery, showing that most types are shared by the two regions, but exhibit marked region-specific differences in proportions and expression programs. Next, we compared macaque and mouse retinal cell types. We found a tight correspondence between mouse and macaque PRs, BCs and ACs, but a divergence in RGC types. Key macaque types and molecular features are also conserved in marmosets and humans. Finally, we used our atlas to determine the cell- and region-specific expression of >190 genes implicated in 7 complex diseases that cause human blindness, demonstrating striking patterns of region- and cell type-specific expression. Overall, our study provides a general framework for using molecularly derived taxonomies of cell types to understand regional and species specializations in the nervous system and other organs.

RESULTS

Comprehensive molecular taxonomy of the primate retina using scRNA-seq

We generated cell atlases of foveal and peripheral regions of *Macaca fascicularis* retina using droplet-based scRNA-seq (Zheng et al., 2017) (STAR Methods), yielding high-quality cell profiles from 92,628 foveal, and 73,053 peripheral cells (Table S1). For foveal samples, cells were dissociated from 0.5-1.5 mm diameter pieces and profiled without further processing. Because ~80% of total cells in peripheral retina are rod PRs, we depleted rods (CD73+) or enrich RGCs (CD90+) from peripheral samples before profiling (Figure S1B, STAR Methods). In parallel, we assembled a retina-specific transcriptome, which substantially improved the mapping of scRNA-seq reads compared to existing references (Figures S1C-E), emphasizing the importance of high-quality tissue-specific transcriptome for scRNA-seq analyses.

To maximize our ability to distinguish cell types consistently, we first grouped cells into the 6 major retinal classes (RGCs, BCs, PRs, ACs, HCs, and non-neuronal cells) based on known class-specific gene signatures (Figures 1C, S1F, Table S1), and then iteratively clustered cells within each class separately, building on our earlier methods (Shekhar et al.,

2016). Altogether, we distinguished 64 foveal (3 PR, 2HC, 12 BC, 27 AC, 16 RGC, 4 non-neuronal), and 71 peripheral (2 PR, 2 HC, 11 BC, 34 AC, 18 RGC, 4 non-neuronal) clusters (Figures 1D-I); all clusters were found in all animals (Figures S1G-I). As noted below, a few clusters contain >1 cell type. Non-neuronal cells were Müller Glia, pericytes, endothelial cells and microglia (Figure 1I). There are no oligodendrocytes in healthy retina, and astrocytes were not recovered in our samples. Using differential expression analysis, we identified molecular markers for each cluster and used this information to assign clusters to individual cell types. We first describe the types, then return to differences between fovea and periphery.

L and M cones are distinguished only by expression of opsin paralogs

Retinas of trichromatic primates, including macaques and humans, contain rods plus three cone types that preferentially detect long (L, red), medium (M, green) and short (S, blue) wavelengths, because of their expression of L, M and S opsins (*OPNILW*, *OPNIMW*, *OPNISW*) respectively. S cones were readily distinguished, whereas M and L cones mapped to a single cluster (Figure 1D). *OPNIMW* and *OPNILW* have ~98% identical coding sequences (Onishi et al., 2002) and are not distinguished in the macaque reference genome. We therefore retrieved exons that distinguish the two opsins from ~70% of the M/L cones, and used them to evaluate M- and L-specific single-nucleotide polymorphisms (SNPs) (Figure S2A-C,E). Of these, >96% from both fovea and periphery exhibited exclusive L or M SNPs (Figure S2B,C,E).

Remarkably, no genes other than *OPNIMW* or *OPNILW* were differentially expressed (DE) between M and L cones (Figures 2A, S2D), supporting a model in which stochastic choice between the neighboring *OPNIMW* and *OPNILW* genes diversifies otherwise identical cones (Wang et al., 1999). In contrast, many genes distinguished M/L from S cones and rods, and S cones from rods (Figures 2B,C and S2F).

Molecular classification and validation of interneuronal (BC, HC, AC) types

BCs, the major excitatory interneurons of the retina, are divided into those that receive predominant input from cones or rods; cone BCs are further divided into those excited (ON BCs) or inhibited (OFF BCs) by light (Euler et al., 2014). Twelve BC types (1 rod, 5 OFF cone and 6 ON cone) have been previously reported in macaque peripheral retina based on morphology and immunohistochemistry (Tsukamoto and Omi, 2015, 2016).

We identified 12 foveal BC and 11 peripheral BC clusters; 10 mapped 1:1 to known types (Table S2), and supervised methods separated another cluster into two known types (Figure S2H). The additional foveal cluster appears to be a previously unreported OFF type, which we call OFFx (Figure S2I). We identified markers for each BC type and validated several of them (Figures 2D,E, S2G,J-O).

HCs and ACs are the main sources of inhibitory input to retinal circuitry. We identified 2 clusters of HCs in both fovea and periphery (Figure 1E), and mapped them to distinct HC subsets in tissue. We also validated *SPP1* (osteopontin) as a pan-HC marker (Figure S5D). We identified 27 foveal and 34 peripheral AC clusters, including both GABAergic and glycinergic subclasses (Figures 1G, S4A, S5E, Table S2). We assigned six clusters to known

mouse types based on their expression of orthologous markers (Figures S4A and S5E). Neuropeptides were expressed in subsets of GABAergic AC types (Table S2).

Markers for midget and parasol RGCs

Morphological studies have distinguished ~18 RGC types in primate retina (Dacey, 2004; Masri et al., 2016). ON and OFF Midget RGCs (MGCs), the smallest and most abundant, comprise 85% of foveal RGCs; larger ON and OFF parasol RGCs (PGCs) comprise ~10%; small bistratified RGCs (SBCs) account for ~5%; and ~11 types with larger dendritic arbors account for the remaining ~1%. MGCs and PGCs also account for a large majority of peripheral RGCs (Dacey, 2004).

We identified 18 peripheral and 16 foveal RGC clusters (Figure 1H). Based on their abundance, we tentatively identified a pair of clusters as ON and OFF MGCs (~85% of RGCs in fovea, 83% in periphery) and a second pair as ON and OFF PGCs (11% in fovea, 4.8% in periphery). Markers for these types include *TBR1* (OFF MGC) *TPBG* (ON MGC) *CHRNA2* (ON PGC) and *CA8* (OFF PGC). *SPP1* and *RBPM2* are expressed by both PGC types and *GUCY1A3* by both MGC types (Figure 2F, S3C-O). We confirmed the identities of these four RGC types by combining FISH with viral labeling to visualize morphology (Figures 2G, S3B,P,Q). Interestingly, RGC somata expressing TFs selective for ON and OFF MGCs and PGCs were segregated into outer and inner halves of GCL, suggesting a molecular basis for the laminar organization of physiologically distinct subclasses (Perry and Silveira, 1988) (Figures 2H and S3P). We were unable to further partition MGCs and PGCs by supervised methods.

Three peripheral RGC clusters expressed *OPN4* (melanopsin), a marker of intrinsically photosensitive RGCs (ipRGCs). We identified markers for the remaining clusters, including *SATB2* as a marker for SBCs (Figures 2I and S3A). To our knowledge, these are the first molecular markers for the major RGC types of the primate retina.

Most peripheral and foveal retinal cell types match 1:1

To relate foveal and peripheral clusters based on gene expression, we devised a multi-class learning framework that associated each foveal cell with a peripheral identity, and then examined the extent to which clusters across the two regions corresponded (STAR Methods).

Most foveal clusters (77%) mapped 1:1 with peripheral clusters including 10/12 BC, 2/2 HC, 20/27 AC and 14/16 RGC clusters (Figures 3A-D). These matches were statistically significant as quantified by the Adjusted Rand Index (ARI), a measure of similarity between data clusterings. Six foveal clusters (RGC clusters fRGC11,14 and AC clusters fGa3,7,13 and fGl8) could be partitioned because they each mapped to two closely related peripheral clusters. Most instances of multi-mapping were for AC clusters, likely due to a biased frequency distribution in the peripheral samples, which were not optimized for AC recovery (Figures S4B).

Quantifying the compositional similarity between the fovea and the periphery suggested that differences decrease from outer to inner retina. Values of the Jensen-Shannon divergence

(JSD) - a distance measure in frequency space ranging from 0 (identical cell type composition) to 1 (completely region-specific) were: PRs (0.21), BCs (0.16), ACs (0.06), and RGCs (0.035) (STAR Methods). There were numerous differences in relative proportions for 1:1 matched types (Figures 3E-H and S4B, Table S2), some consistent with previous reports (Grunert and Martin, 1991; Wässle et al., 1994). Six RGC types were >4-fold enriched in foveal compared to peripheral samples (Figure 3G), 6 GABAergic AC types appeared to be periphery-specific (Figure 3C), and the OFFx BC type appeared to be fovea-specific (Figure S2I). Taken together, these results suggest that foveal and peripheral neuronal circuitry draw upon a similar but not identical “parts list” of cell types.

Programs for phototransduction and GABAergic neurotransmission differ between matching types in the fovea and periphery

We next asked whether corresponding foveal and peripheral cell types exhibited DE genes (Figures 4, S4). We found a median of 17±8 genes significantly DE between corresponding types (Table S3), with more differences for RGCs and non-neuronal cells (median 130 genes/type) than for interneurons (HCs, BCs and ACs; median 13 genes/type) (Figure 4A).

One intriguing difference is in *GNGT1*, which encodes the gamma subunit of transducin, a G protein involved in phototransduction. Light responses are slower in macaque foveal than peripheral cones (Sinha et al., 2017), and homozygous deletion of *Gngt1* accelerates rod responses in mice (Kolesnikov et al., 2011). *GNGT1* is selectively expressed by rod PRs in mice and macaques (Figure 4B), but is also expressed at higher levels in foveal than peripheral cones, as verified by FISH and immunohistochemistry (Figure 4C and see below). It is thus tempting to speculate that the enrichment of *GNGT1* in macaque foveal cones contributes to the kinetic differences, possibly by allowing transducin to compete more effectively with other molecules that deactivate the visual pigment (Doan et al., 2009). Conversely, a cone-specific phototransduction component, *GNB3* is also expressed at substantial levels by foveal rods (Figure 4B).

Another prominent difference is that several genes involved in GABAergic neurotransmission are expressed at higher levels in peripheral types than in their foveal counterparts. They include the GABA_A receptor subunit *GABRB2* in PRs (Figures 4B,C), the GABA synthetic enzyme *GAD2* in rod BCs (Figure 4D) (Lassova et al., 2010) and several GABA receptor subunits in multiple RGC types (Figure 4F-H and S4C). Thus, the paucity of inhibitory drive to foveal MGCs observed physiologically (Sinha et al., 2017) may reflect both the absence of several GABAergic AC types from fovea (Figure 3C) and decreased expression of GABAergic components in shared cell types.

We also observed >100 DE genes between foveal and peripheral MGs, including *SPP1* enriched in peripheral MGs and *CYP26A1* in foveal MGs (Figures 4I-K, S4D). These differences provide potential substrates for differences between their interactions with neurons in the two regions (Bringmann et al., 2018).

Tight molecular correspondence between macaque and mouse BC and AC types

While the conservation of genes across species has been extensively studied, we know less about the extent to which cell types are conserved. scRNA-seq enables systematic

comparison of cell atlases across species by leveraging gene orthology (Marioni and Arendt, 2017). To exploit this opportunity, we adapted the multi-class classification framework used for fovea-periphery comparisons to compare macaque cells with those previously collected from mice (Macosko et al., 2015; Shekhar et al., 2016).

As expected from the conserved retinal architecture, there was a 1:1 match between cell classes in the two species (Figure S5A). In addition, there was also a tight correspondence between mouse and macaque PR, HC, BC and AC types. Mice are dichromats, with only M and S cone opsins, and many cones express both opsins (Euler et al., 2014). Mouse M and S cones most closely resembled macaque M/L and S cones, respectively (Figures 5A, S5B). Mouse HCs were more transcriptionally similar to macaque H1, which they resemble morphologically, than to H2 (Wassle et al., 2000) (Figures 5B, S5C,D, Table S2). For ACs, there was a striking correspondence between macaque and mouse, based on comparison to our previously reported (but underpowered) estimate of 21 molecularly distinct mouse AC clusters (Macosko et al., 2015) (Figure 5C), with clear macaque equivalents of several well-studied mouse types (Figures 5D, S5E and Table S2). A more comprehensive ongoing study has revealed >60 mouse AC types, and increased the 1:1 matches with macaque (W.Y., K.S., A.R., and J.R.S. in preparation).

For BCs, 9 peripheral macaque types mapped preferentially to mouse types (Figure 5E; 8 were 1:1), and expression patterns of type-enriched orthologues were conserved between the two species (Figure 5F). As noted below, we were unable to identify clear murine equivalents of the most abundant RGC types, MGCs and PGCs. However, the BCs that selectively innervate MGCs (FMB and IMB) correspond to mouse BC1 and 7; and the BCs that provide strongest innervation to PGCs (DB2, 3a and 4) (Tsukamoto and Omi, 2015, 2016) correspond 1:1 to mouse BC4, 3a and 5a. It will be interesting to ask whether postsynaptic targets of these BC types share features with MGCs and PGCs.

Divergence in expression programs between macaque and mouse RGC types

In contrast to BCs and ACs, mouse and monkey RGC types differ greatly in both number (>40 in mice [Baden et al., 2016; Bae et al., 2018] vs. <20 in macaque) and distribution (no single mouse type accounts for >10% of all RGCs whereas MGCs account for >80% of macaque RGCs). Moreover, using a recently published scRNA-seq atlas of early post-natal mouse RGCs (Rheume et al., 2018), we found few clear matches between the species (ARI = 0.16), one being the evolutionarily ancient melanopsin-positive intrinsically photosensitive RGCs (Figures 5G,I); macaque RGC types exhibit a similar lack of correspondence with adult mouse RGC types (K.S., W.Y., A.R. and J.R.S. *in preparation*). Although limited sampling might reduce our ability to detect conservation, we found no clear mouse equivalents of the abundant macaque MGCs (~35,000 cells).

Nonetheless, we found similar patterns of TF expression across macaque and mouse RGC clusters. For example, *Tbr1*, *Eomes*, *Satb2*, *Tbx20* and *Foxp2*, several of which regulate type-specific features of mouse RGCs (Liu et al., 2018; Mao et al., 2014; Peng et al., 2017), are also expressed selectively by restricted macaque types (Figure 5H). Among them, *Satb2* is expressed by bistratified RGCs in both macaque (SBCs; Figure 2I) and mice (ooDSGCs; Peng et al., 2017). Moreover, patterns of TF co-expression were similar in both species

(Figure S5F), suggesting a model wherein subfamilies of RGC types are specified by TF codes, but with substantial divergence of the regulatory targets of those codes.

Conservation of cell types and markers across primates

A main value of studies in non-human primates is that their properties are likely to be shared with humans. To ask whether this is true of retina, we performed preliminary analyses on humans as well as a second non-human primate, the common marmoset (*Callithrix jacchus*), which is increasingly used because of its small size and genetic accessibility.

We found strong conservation of key molecular and cellular features across macaque, human and marmoset retinas. For example, markers of MGCs, PGCs, S cones and HC types were conserved in marmoset (Figures 6A, S6A,B) and the laminar segregation of foveal OFF and ON RGCs was apparent in both marmoset and human (Figures 6B, S6D). Distinctions between corresponding peripheral and foveal cell types were also conserved, such as the enrichment of *GNGT1* in foveal cones (Figure S6C) and enrichment of *CYP26A1* and *SPPI* in foveal and peripheral MGs respectively (Figures 6C,D and S6E, 7E). Moreover, we profiled 2,383 BCs from human peripheral retina and found close correspondence (ARI = 0.74) between human and macaque types (Figures 6E,F, S6F-H). Together with previous morphological studies (Haverkamp et al., 2003; Liao et al., 2016), these results reveal strong conservation of cell types among primates.

Cell type specific expression of human retinal disease genes

Finally, encouraged by the correspondence between human and macaque cell types, we assessed the expression of genes that have been implicated in 7 diseases associated with disabling vision loss: retinitis pigmentosa, cone-rod dystrophy, diabetic macular edema and retinopathy, congenital stationary night blindness, primary open angle glaucoma, primary angle closure glaucoma, and age-related macular degeneration. We focused on ~200 genes for which association has been demonstrated by loss- or gain-of-function mutations or by Genome-Wide Association Studies (GWAS) studies (Farrar et al., 2017; Fritsche et al., 2016; Graham et al., 2018; Wiggs and Pasquale, 2017; Zeitz et al., 2015).

For each gene we calculated an enrichment score for each macaque foveal and peripheral cell class. Aggregating these scores by disease groups (Figure 7A) shows that, in general, expression was highest in cell classes primarily affected by the diseases: PRs for retinitis pigmentosa and rod-cone dystrophy; PRs and BCs for congenital stationary night blindness; RGCs for glaucoma; and non-neuronal cells for diabetic macular edema and retinopathy. As expected, many genes implicated in macular degeneration were expressed in PRs and non-neuronal cells, but several were also highly expressed in other cell classes, especially ACs, a class whose role in disease etiology and progression remains to be explored.

Patterns of expression for several genes are noteworthy. At least three genes associated with diabetic macular edema which, as the name implies, selectively affects the macula, were expressed at higher levels in certain foveal non-neuronal types than their peripheral counterparts, such as *PDGFB* and *EDNI* in endothelial cells (Figure 7C) (Graham et al., 2018). Likewise, the macular degeneration susceptibility gene *VTN* (Fritsche et al., 2016) was enriched in foveal compared to peripheral rods and cones (Figure S7). *HTRA1*, a major

susceptibility gene for macular degeneration (Fritsche et al., 2016), which has been implicated in dysfunction of retinal pigment epithelium is also expressed at high levels in macaque but not in mouse HCs. Glaucoma associated genes *CYP1B1* and *CYP26A1* were enriched in a GABAergic AC type and foveal MGs, respectively (Figure 7B,E and S7). Several other glaucoma-associated genes were enriched in specific RGC subsets: *MAP3K1* by MGCs and PGCs; *SIX6* by MGCs; and *GAS7* in multiple non-MGC and PGC types (Figure 7D). These region- and cell type-selective expression patterns can inform future analyses of genetic variants associated with disease risk. Interestingly, expression patterns of many human retinal-disease genes were not conserved in mice, highlighting the limitations of using a mouse atlas as a reference (Figure S7).

DISCUSSION

We used scRNA-seq to generate a cell atlas of the primate retina, and used it as a foundation to address several questions about retinal structure and function. First, to explore regional differences within a tissue, we compared the cellular compositions of macaque fovea and peripheral retina. Second, to assess evolutionary specializations across species, we compared macaque (which have a fovea) and mice (which do not), and also asked whether foveal specializations in macaque are conserved in two other primates, human and marmoset. Finally, we analyzed the expression of nearly 200 genes that have been associated with blinding diseases.

Primate retinal cell atlas

We detected >70 cell types in peripheral retina and >65 in fovea. Although extremely rare retinal cell types may have gone undetected, we believe our atlas is reasonably complete. The number of cells needed to generate a comprehensive atlas depends on multiple factors, including the number and the relative frequency of types and the extent to which types differ from each other (Shekhar et al., 2016). We analyzed >165,000 cells and, for peripheral retina, depleted abundant rods from some samples and enriched RGCs in others. These strategies enabled us to detect very rare cell types, such as ipRGCs, which comprise <0.002% of all retinal cells (Liao et al., 2016). Complete compendia of cell types from comparably complex tissues will likely require similar cell numbers and/or enrichment procedures.

One product of the atlas is a set of molecular markers that can be used to mark specific cell types. Among these, the ON and OFF MGCs and PGCs are particularly noteworthy. These four types comprise ~95% of foveal and ~85% of peripheral RGCs; they therefore transmit the vast majority of information from eye to brain. Yet, to date, there have been no means to identify them other than by laborious and low throughput physiological or morphological methods. The availability of markers now enables a variety of studies on these indispensable cell types.

We also explored differences between L and M cones. These cell types transmit different spectral information but the question of how they do so has been controversial. If M and L cones connect selectively to different postsynaptic cell types, as some studies suggest (Lee et al., 2012), one would expect them to express different recognition molecules, implying

transcriptional differences that extend beyond the opsin genes. If wiring is nonselective, as indicated by other studies (Wool et al., 2018), no non-opsin expression differences would be needed, implying that our ability to distinguish green from red is experience-dependent rather than innate. Our results are consistent with the latter view. However, it is possible that subtle distinctions between L and M cones might have gone undetected, or critical differences might be restricted to early stages, when cones form connections (Hoshino et al., 2017).

Fovea and periphery

The fovea mediates most high-acuity vision and much chromatic vision in primates. Indeed, one view of the peripheral retina is that at moderate to high light levels, its main job is to aid in directing eye movements to focus salient images on the fovea. Numerous physiological and morphological differences between fovea and periphery have been documented (see Introduction) but lack of molecular information has made it difficult to elucidate the extent to which differences stem from cell types unique to one region, from altered compositions across the same set of types, or from region-specific molecular programs within shared types.

Our results suggest that all three factors are involved, but to different extents. A few cell types are unique to or highly enriched in each region, and these may underlie certain functional differences between them. The OFFx bipolar is particularly interesting in this regard, and the absence of some foveal GABAergic AC types may contribute to the low level of foveal inhibition (Sinha et al., 2017). Nevertheless, our results suggest that the dramatic specializations of foveal circuitry and function arise mostly from regional differences in proportion and gene expression within shared types rather than from region-specific types. The differences in proportions (as measured by JSD) appear to be higher at the sensory and preprocessing layers (PRs and BCs) than for downstream processors and feature detectors (ACs and RGCs). Some of the differentially expressed genes, such as components of phototransduction and neurotransmitter pathways, may underlie documented differences in visual responses mediated by the fovea and periphery (Sinha et al., 2017).

Macaques and Mice

It has been clear for over a century that the major retinal cell classes are conserved across vertebrates (Cajal 1892). Far less is known about the conservation of types within classes. Comparison of macaque and mouse allowed us to address this issue. Conservation was stronger for types within outer retinal classes (PRs, BCs and ACs) than for RGCs. Most notably, MGCs, the major RGC types in primates, lack clear counterparts in mouse. Similarly, the molecular distinctions between the fovea and the periphery are higher for RGCs than for PRs and interneurons (Fig. 4A). These patterns suggest that the outer retina may comprise a conserved set of information processors, with adaptation to species- and region-specific visual processing requirements beginning at the level of RGCs, and then continuing centrally.

This is not to say that macaque and mouse RGC types are unrelated. There may be an underlying shared gene regulatory program that specifies RGC diversity, based on conserved

patterns of type-specific TF expression and co-expression. These factors may specify groups that diversify differentially in the two species, raising the possibility that cross-species differences in homologous regions can be explained by rewiring of downstream programs specified by a conserved set of TFs.

In addition to the paucity of 1:1 matches, the number of RGC types is far lower in macaques than in mice. This difference may seem paradoxical, given the greater visual capabilities of primates. One possible explanation is that primates may rely less than mice on a broad array of retinal feature detectors (e.g., direction-selective RGCs) and more on complex processing in the cortex. Thus, primate RGCs, especially MGCs, may be designed to transmit a relatively unprocessed image with little loss of information to the cortex, where more sophisticated and flexible computations are performed. Presumably what is lost in speed is more than compensated by finesse.

Retinal diseases

Many genes have been implicated in the pathogenesis of retinal diseases, some based on rare but highly penetrant mutations and others as common variants that confer susceptibility (Farrar et al., 2017; Fritsche et al., 2016; Graham et al., 2018; Wiggs and Pasquale, 2017; Zeitz et al., 2015). We assessed expression of nearly 200 such genes and found that many are selectively expressed in particular retinal cell classes, in particular types within classes, and in foveal or peripheral cohorts of particular types. For example, *PDGFB* and *EDNI*, implicated in diabetic macular edema (Graham et al., 2018), are expressed at significantly higher levels by foveal than peripheral endothelial cells, suggesting a basis for the increased susceptibility of the macula in this disease. Interestingly, several of the patterns we document in macaque are not conserved in mice, emphasizing the value of using primate models for investigating human retinal dysfunction. Together, these results provide both technical and conceptual foundations for investigations aiming to understand cellular mechanisms underlying blinding retinal diseases.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Joshua R. Sanes (sanesj@mcb.harvard.edu).

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Tissue Procurement—Non-human primate and human retinas were obtained and used in accordance with the guidelines for the care and use of animals and human subjects at Harvard University and Boston Children's Hospital, and supplying institutions. All the procedures on non-human primates were approved by the Institutional Animal Care and Use Committees. Acquisition and use of human tissue was approved by the Human Study Subject Committees (DFCI Protocol Number: 13-416 and MEE - NHSR Protocol Number 18-034H).

Eyes from male macaques (*Macaca fascicularis*, 3-9 years of age) were kindly provided by institutions including Biomere and Massachusetts General Hospital. Eyes were collected either pre-mortem under deep anesthesia or 45 min post-mortem. In some cases, whole globes were immediately placed in ice-cold Ames solution (Sigma-Aldrich; equilibrated with 95% O₂/5% CO₂ for all use), where they were stored before experimentation. In others, a rapid hemisection was performed to remove the vitreous and the anterior chamber, and the posterior eyecup was immersed in room-temperature Ames. Retinas were then dissected free and stored in ice-cold Ames solution. Experiments commenced within 8 hours. For access to macaque tissue, including the samples used in this study, we thank V. Belov, C. Cetrulo, D. Guberski, A. Hall, P. Kovalenko, A. LaRochelle, J. Madsen, M. Nedelman, M. Papisov, and S. Smith.

Eyes from male and female common marmosets (*Callithrix jacchus*, 2-10 years of age) were generously provided by the McGovern Institute for Brain Research (Massachusetts Institute of Technology). Animals were perfused with 4% paraformaldehyde (PFA) under deep anesthesia and post-fixed for 1 hr in 4% PFA. Eyes were then collected and stored in ice-cold Ames solution. We are grateful to G. Feng, Q. Zhang and C. Wu for access to this tissue.

Macaque and marmoset eyes were collected from animals that had reached the end of unrelated studies at supplying institutions. No ocular or visual abnormalities were noted. Data presented in this manuscript did not covary with any treatment that had been applied to the animals.

The human eye used for sequencing and *in situ* hybridization was collected ~6 hours postmortem from a 74 year-old male. The whole globe was immediately transported back to the lab in a humid chamber. Hemisection was performed to remove the anterior chamber and the posterior pole was recovered in Ames equilibrated with 95% O₂/5% CO₂ before further dissection and dissociation. The donor was confirmed to have no history or clinical evidence of ocular disease or intraocular surgery. We are grateful to Dr. Juric and the Rapid Autopsy Program, Susan Eid Tumor Heterogeneity Initiative, Massachusetts General Hospital for expeditious access to this material, enabling recovery of high quality RNA for sequencing.

Other human eyes used for *in situ* hybridization were provided by the Lions VisionGift (Portland, OR) and the National Disease Research Interchange (NDRI; Philadelphia, PA). They were collected <11 hr postmortem from a 28 year-old male and a 69 year-old female, hemisected, and fixed overnight in ice-cold 4% PFA following removal of the cornea. No ocular disease was reported in these donors.

METHOD DETAILS

RNA-sequencing

Single Cell Isolation: 0.5-1.5 mm diameter foveal tissues centering on the foveal pit were dissected out from four macaque eyes. Foveal samples were digested with 200 units papain (Worthington, LS003126). Foveae M1 and M2 were digested at 37°C; foveae M3 and M4 were digested at 22°C. Following digestion, retinas were dissociated and triturated into single cell suspensions with 0.04% bovine serum albumin (BSA) in Ames solution.

Peripheral retinal pieces were dissected and pooled from all quadrants of the retina. Single cell suspensions were dissociated at 37°C as described for fovea. Dissociated cells were incubated with CD90 microbeads (Miltenyi Biotec, 130-096-253; 1 µl per 10⁷ cells) to enrich RGCs or with anti-CD73 (BD Biosciences, clone AD2; 5 µl per 10⁷ cells) followed by anti-mouse IgG1 microbeads (Miltenyi Biotec, 130-047-102; 10 µl per 10⁷ cells) to deplete rods. Incubations were at room temperature for 10 min. CD90 positive cells or CD73 negative cells were selected via large cell columns through a MiniMACS Separator (Miltenyi Biotec). Single cell suspensions were diluted at a concentration of 500-1800 cells/µL in 0.04% BSA/Ames for loading into 10× Chromium Single Cell A Chips. Human cells were dissociated and treated with anti-CD73 as above.

Droplet-Based scRNA-seq: Single cell libraries were prepared using the Chromium 3' v2 platform (10× Genomics, Pleasanton, CA) following the manufacturer's protocol. Briefly, single cells were partitioned into Gel beads in EMulsion (GEMs) in the GemCode instrument followed by cell lysis and barcoded reverse transcription of RNA, amplification, shearing and 5' adaptor and sample index attachment. On average, approximately 10,000 single cells were loaded on each channel and approximately 6,000 cells were recovered. Libraries were sequenced on the Illumina HiSeq 2500 (Paired end reads: Read 1, 26bp, Read 2, 98bp).

Histological methods

Viral Transduction: Macaque or marmoset retinas were divided into four quadrants according to cardinal axes with the macula in the temporal region and mounted on Millicell culture inserts (Millipore) with the ganglion cell layer facing up. Retinas were infected with 5 µl of AAV retrograde virus encoding GFP directed by the CAG promoter (Addgene) with a titer ~10¹². Methods were adapted from (Johnson and Martin, 2008; Meyer-Franke et al., 1995). Transduced retinas were cultured in Neurobasal A medium (Fisher 10888-022) supplemented with 0.1 mg/ml penicillin/streptomycin (Sigma G4333), 1 mM L-glutamine (Sigma G7513), B27 (Fisher 17504-044), 0.005 mM forskolin (Sigma F6886), 1 mM sodium pyruvate (Fisher 11360-070), 25 ng/ml BDNF (PeproTech 450-02), and 10 ng/ml CNTF (PeproTech 450-13) for four days. Retinas were then fixed in 4% PFA in phosphate-buffered saline (PBS) for 1 hr at 4 °C and separated from the culture insert after fixation.

Biolistic Transfection: The biolistics procedure was adapted from a published protocol (Masri et al., 2016). Briefly, gold particles (1.6 µm diameter; 12.5 mg; Bio-Rad) were coated with pCMV-GFP DNA plasmid (25 µg). The particles were delivered to retinal cells in whole-mount retinas with GCL facing up using a Helios gene gun (Bio-Rad). The retinal explants were cultured for 3 days followed by fixation as described for virus transduced retinas.

Fluorescent In Situ Hybridization (FISH): Eyes were fixed in 4% PFA. Marmoset and human eyes were fixed by perfusion and immersion respectively as described under "Tissue Procurement." For macaque eyes, slits or windows were made in the cornea of eyes collected post-mortem, and the globe was immersion-fixed overnight in ice-cold 4% PFA. Probe templates were generated using cDNA derived from adult macaque, marmoset, or

human retinas following RNA extraction and reverse transcription with AzuraQuant™ cDNA Synthesis Kit (Azura, AZ-1995). Antisense probes were generated by PCR using a reverse primer with a T7 sequence adaptor to permit *in vitro* transcription (see Table S4 for primer sequences). DIG rUTP (Roche, 11277073910), DNP rUTP (Perkin Elmer, NEL555001EA), and Fluorescein rUTP (Roche, 11685619910) were used for synthesizing probes for single, double or triple fluorescent *in situ* hybridization (FISH) experiments. Retinas fixed as above were rinsed with PBS, immersed with 30% sucrose overnight at 4 °C, embedded in Tissue Freezing Medium (EMS) and cryosectioned at 20 µm. Details of FISH were described in ref. 12. Briefly, sections were mounted on Superfrost slides (Thermo Scientific), treated with 1.5 µg/ml of proteinase K (NEB, P8107S), and then post-fixed and treated with acetic anhydride for deacetylation. Probe detection was performed with anti-DIG HRP (1:1000), anti-DNP HRP (1:500), anti-Fluorescein (1:1000) followed by tyramide amplification. Detection of protein epitopes was performed following probe detection. Antibodies were diluted in block consisting of 3% donkey serum (Jackson, 017-000-121), and 0.3% Triton-X in PBS at concentrations of 1:500 (anti-GFP), 1:1000 (anti-Calbindin), 1:2000 (anti-PKCa), and 1:100 (anti-CD15) (See below for antibody information).

Immunohistochemistry: Tissue was fixed as described above. Immunostaining was performed as described in refs. 12 and 31. Antibodies used were as follows: chick and rabbit anti-GFP (1:500, Abcam; 1:2000, Millipore); mouse anti-Satb2(1:1000, Abcam); goat anti anti-CHX10 (1:300, Santa Cruz); rabbit anti-TFAP2B (1:1000, cell signaling); goat anti-Choline Acetyltransferase (1:500, Millipore); goat anti-VACHT (1:1000, Millipore); rabbit and guinea pig anti-RBPMS (1:1000, Abcam; 1:5000, PhosphoSolutions); rabbit anti-Calbindin (1:2000, Swant); mouse anti-Calretinin (1:5000, Millipore); mouse anti-Gad65/67 (1:1000, Millipore); mouse and rabbit anti-PKCa (1:2000, Abcam; 1:2000, sigma); rabbit anti-Secretagogin (1:10,000; BioVendor); mouse anti-Human CD15 (1:100, BD); rabbit anti-GNGT1 (1:1000, gift from Dr. N. Gautam), mouse anti-7G6 (1:1000, gift from Dr. Peter MacLeish), and chick anti-OPN1M/LW (1:2000, gift from Dr. Jeremy Nathans). Nuclei were labeled with DAPI (1:1000, Invitrogen). Secondary antibodies were conjugated to Alexa Fluor 488, 568, and 647 (Invitrogen) and used at 1:1000. Sections and whole mounts were mounted in ProLong Gold Antifade (Invitrogen).

Image Acquisition, Processing and Analysis: Images were acquired on Zeiss LSM 710 confocal microscopes with 405, 488-515, 568, and 647 lasers, processed using Zeiss ZEN software suites, and analyzed using ImageJ (NIH). Images were acquired with 16×, 40× or 63× oil lens at the resolution of 1024×1024 pixels, a step size of 0.5-15µm, and 90µm pinhole size. ImageJ (NIH) software was used to generate maximum intensity projections and neuronal dendrites were reconstructed with ImageJ plugin simple neurite tracer. Adobe Photoshop CC was used for adjustments to brightness and contrast.

Computational Methods

Assembling a Retina-specific Transcriptome for *M. fascicularis*: We obtained high quality total RNA from macaque retinal tissue (RNA Integrity Number (RIN) score 9.8), and prepared strand-specific libraries using the TruSeq strand-specific Total RNA kit (Illumina Inc.), which was sequenced on the NextSeq 500 system to obtain 45 million 100bp paired

end reads. We used StringTie (v1.3.3) (Pertea et al., 2016) to assemble a genome-guided transcriptome, using an available NCBI transcriptome for *M. fascicularis*¹ (annotation release 101) as an initial guide. Briefly, we mapped RNA-seq reads onto the existing transcriptome in a strand-specific manner using the Hisat2 software (with command line options --dta --rna-strandedness RF), following published guidelines (Pertea et al., 2016). Next, we used StringTie (with command line option --rf) to assemble a new transcriptome based on TruSeq reads. We then reran StringTie (with the command line option merge) to obtain an updated transcriptome annotation, which contained modifications of transcript body definitions from the existing NCBI reference as well as novel transcripts supported by the TruSeq reads. While the modified transcripts retained gene names from the original annotation, the novel transcripts were initially named according to Stringtie's naming convention (e.g. MSTRG.5141). We call this the NCBI + TruSeq reference in Figure S1. There were a few instances where sense and antisense transcripts were predicted for some genes; in these cases we added suffixes "_p" and "_n" against the corresponding gene name to indicate this fact (e.g. *FEZF1_p* and *FEZF1_n*).

To facilitate transcriptional mapping of macaque types to mouse and human types, we used BLAT (Kent, 2002) to associate each macaque transcript with its closest ortholog (reciprocal best matches) in the NCBI mouse (version GRCm38) and human (version GRCh37) transcriptomes, respectively. Encouragingly, this automated procedure matched transcript gene names with similar letter codes (e.g. macaque *TRPM1* with mouse *Trpm1* and human *TRPM1*). We use this mapping to associate each reconstructed transcript with the closest human gene by sequence, in cases where such homology was strong. In some cases, we also found that certain tenuously named loci in the NCBI reference could be associated with homologous mouse and human gene names (e.g. macaque *LOC101864869* could be associated with mouse *Rps27a*, *LOC102132859* could be associated with mouse *Cyp26a1* etc). For many loci, however, we were unable to associate a human gene and for these cases, we retained the naming convention assigned by StringTie (e.g. MSTRG.5141).

Pre-Processing Of 3' Droplet-based scRNA-Seq Data: Sample demultiplexing, alignment to the NCBI+TruSeq reference, quantification and initial quality control (QC) was performed using the Cell Ranger software (version 2.1.0, 10x Genomics) for each sample (*i.e.*, 10x channel) separately. We used the option "--force-cells 6000" in Cell Ranger count to obtain a 36,162 genes × 6,000 cells count matrix each sample to deliberately extract a larger number of cell barcodes in the data, as we found that the automatic estimate of Cell Ranger was too conservative, and was unfavorable for small cell-classes like BCs. Here, 6,000 represented a "loose" upper bound on the number of cells that could be recovered given the density of each the cell suspension loaded onto every channel per the manufacturer's estimates. We grouped the count matrices separately for the foveal and the peripheral samples to generate a consolidated matrix for each region. Only cells which expressed >500 genes were retained for further analysis. Further pruning of cells (low quality cells, doublets) was done for each class separately (see below). Library size, as quantified by the number of genes detected per cell on average, was highly cell-class

¹<https://www.ncbi.nlm.nih.gov/genome/776>

dependent. Values by class, from highest to lowest, were: RGCs, 4000 genes/cell; ACS, 1500 genes/cell; cones, 1400 genes/cell; non-neuronal cells, 1200 genes/cell; HCs, 1000 genes/cell; rods, 900 genes/cell; and BCs, 750 genes/cell).

Separation of Major Cell Classes: Both the foveal and the peripheral datasets consisted of multiple biological replicates (Figs. S1G-I). Moreover, depending on the enrichment method, each sample contained widely different distributions of the main cell classes. For example, the foveal samples typically comprised of ~30% RGCs, ~28% BCs, ~8% ACs (Figure S1H), the peripheral CD90+ samples contained ~28% RGCs, 0% BCs, ~50% ACs, and the peripheral CD73- samples contained <1% RGCs, ~30% BCs and ~10% ACs (Figure S1I). Moreover, we anticipated that the differences in proportions also extends to types within a class (Figure 3), and that foveal and peripheral types likely had molecular differences (Figure 4). This made conventional batch correction using methods such as Surrogate Variable Analysis and ComBat difficult, because they make strict distributional assumptions that are violated here. Furthermore, covariates that can be used to counteract these assumptions are not known a priori. We also attempted to perform global batch correction using recently published methods that use Canonical Correlation Analysis (CCA; (Butler et al., 2018)) or Mutual Nearest Neighbour (MNN;(Haghverdi et al., 2018)) matching, but in many cases these did not completely remove batch effects, collapsed distinct types with known molecular markers, or appeared to have varying impact on different cell classes.

To make batch correction more tractable and to avoid extensive biases in the initial stages for the reasons stated above, we first analyzed the foveal and the peripheral datasets separately to stratify the cells by their class. We followed our earlier PCA + Louvain-Jaccard graph-clustering pipeline (Shekhar et al., 2016) without batch correction to obtain a set of transcriptionally distinct clusters. At this stage, we deliberately set parameters to “overcluster” the data to avoid combining distinct cell classes. Next, we scored each cell for gene signatures of well-known retinal cell classes - Rods, Cones, Bipolar Cells (BC), Horizontal Cells (HC), Amacrine Cells (AC), Retinal Ganglion Cells (RGC), and non-neuronal cells such as Müller Glia (MG) and others (Table S1). Classes with shared markers often scored similarly, but with one higher than the other (e.g. *TFAP2A/2B* and *ONECUT1/2* are expressed in ACs and HCs, with *TFAP2A/2B* being higher in the former and *ONECUT1/2* being higher in the latter). In such cases, we combined these classes and analyzed them as a group in downstream pipelines. Thus, we separated the foveal and peripheral datasets into 5 groups comprising - (i) Rods and Cones, (ii) Amacrine Cells (ACs) and Horizontal Cells (HCs), (iii) Bipolar Cells (BCs), (iv) Retinal Ganglion Cells (RGCs), and (v) Others consisting predominantly of Müller Glia, but also pericytes, endothelial cells and microglia (Figure. 1C). Astrocytes are absent from the fovea and are largely confined to the optic fiber layer in the periphery (23). None were recovered from either region in our samples. Encouragingly, cells of the same class were more highly correlated with each other than cells of other classes, supporting this separation (Figure S1F).

Although clusters at this stage exhibited batch effects, ~90% of clusters could be unequivocally assigned to a class. The other clusters comprised low quality cells and “doublets.” These were often characterized by overall low gene count that was tightly

distributed around the minimum cutoff value (~500-700 genes/cell), the mixed expression of more than one class- specific scores albeit at lower levels than “pure” clusters, and the absence of expression of non- class markers that were specific. Moreover, these cells typically clustered proximally on the dendrogram or tSNE visualization to another large, bonafide cluster, as observed previously (Shekhar et al., 2016). To avoid discarding cells erroneously, we analyzed these clusters separately (not shown), and confirmed that they were doublets and low-quality cells. Because the excluded cells were heavily biased towards low gene-count cells, we could have excluded them *ab initio* by setting a more stringent threshold, but this would have risked excluding genuinely small, but intact single- cell libraries.

Dimensionality Reduction, Clustering and Visualization For Each Foveal and

Peripheral Group: Following the initial clustering step, we analyzed the 6 major classes of cells - RGCs, BCs, ACs, HCs, PRs, and non-neuronal cells for fovea and periphery separately (Figures 1D-I). HCs and ACs were separated for this analysis. We employed the following computational steps to cluster each of these groups as described below.

1. Normalization: Expression values E_{ij} for gene i in cell j were calculated following (Shekhar et al., 2016). Briefly, the Cell Ranger reported UMI (i.e. transcript) count value for each gene i in each cell j was divided by the sum of the total UMI counts in cell j to normalize for differences in library size, and then multiplied by M , the median UMI counts for all cells within the group, resulting in Transcripts-per-median ($TPM_{i,j}$) values. E_{ij} was then calculated as $\log(TPM_{i,j} + 1)$

2. Identification of highly variable genes (HVGs): We first calculated the mean (μ) and coefficient of variation (CV) of transcript counts for each gene in the data. We then computed for each gene, the “excess CV” (or eCV) by subtracting the observed CV from a null model of CV vs μ . This null model is based on a Poisson-Gamma mixture, which was shown to accurately model the null CV vs μ relationships for data containing UMIs for a wide range of 3’-biased protocols (Pandey et al., 2018). We calculated the mean and standard deviation (sd) of the eCV values, and selected genes (HVGs) that had $eCV > \text{mean} + 0.7 * \text{sd}$. This typically selected 700-2500 HVGs for each group.

3. Batch correction: We restricted the expression matrix E_{ij} to HVGs and used a linear regression model (adapted from the source code of the ‘RegressOut’ function of the R package ‘Seurat’) to remove correlations of expression values associated with three covariates – the animal of origin (Figure S1G), the total number of genes observed per cell, and the expression strength of ribosomal genes within each cell. The choice of covariates was guided by running initial tests on PRs, BCs and RGCs where some prior knowledge of the underlying types exists. We tested different combinations of covariates, and settled on a combination that led to the alignment of known types across replicates (e.g., cones or IMBs), and used this scheme consistently for all cell group analyses. Here, we also tested the CCA-based (Butler et al., 2018) and MNN-based (Haghverdi et al., 2018) batch correction strategies, but found that the performance of these methods depended on the choice of certain tunable parameters, which were different for different groups. The linear regression

model, on the other hand, did not contain additional parameters outside of the choice of covariates, and was computationally much faster than the other two methods. In a few cases, some residual batch effects still remained, which we corrected for in a supervised manner (see below). For BCs and RGCs, two groups extensively validated in this study, we additionally confirmed the absence of any artifacts by re-clustering single batches separately and verifying that all the abundant clusters were preserved.

We refer to the corrected expression matrix as \tilde{E} . For each of the HVGs, we compute the Pearson correlation coefficient between the values in E and those in \tilde{E} . Encouragingly, most genes had a similar correlation coefficient ($r > 0.97$), but ~30 genes had a significant reduction in correlation ($r < 0.6$). We flagged these “batch specific genes” in our differential-expression analysis of clusters in 5 (see below).

4. Principal component analysis and 2D visualization: We restricted the corrected expression matrix \tilde{E} to HVGs, and values were centered and scaled before input to Principal Component Analysis (PCA), which was implemented using the R package ‘irlba’. Statistically significant PCs were estimated using the Tracy-Widom distribution (Shekhar et al., 2016), which we found to be an excellent approximation to exact values computed using the more exact, but computationally intensive, bootstrap permutation test (Shekhar et al., 2016). For visualization purposes only, dimensionality was further reduced to 2D using t-distributed stochastic neighbor embedding (t-SNE). We used a recently developed rapidly accelerated implementation (Linderman, 2017) of t-SNE setting the “perplexity” parameter to 30. The resulting t-SNE cell coordinates were used to generate the visualizations in Figure 1D-I.

5. Graph-based clustering: To partition the data into clusters of transcriptionally related cells, we used unsupervised clustering based on the Louvain algorithm with the Jaccard correction (Shekhar et al., 2016). We first constructed a k -nearest neighbor graph (kNN) using the Euclidean distance metric between pairs of cells in the reduced dimension space of the significant PCs, as implemented in the R package ‘RANN’. $k=30$ was used for all groups, except peripheral HCs where $k=15$ was used because of the low cell count.

6. Cluster refinement: As the Louvain method suffers from a well-known “resolution” limit (Fortunato and Barthelemy, 2007), we refined the set of initial clusters as well as corrected for potential overclustering due to residual batch effects in three steps as follows. *Step 1* and *Step 2* were performed for each of the initial clusters in (5) separately. In *Step 1*, each cluster was sub-clustered using the more sensitive Infomap community detection algorithm (Rosvall and Bergstrom, 2008) on a new k -nearest neighbor graph ($k = 5 \cdot \log(\text{cluster size})$). However, this often ended up overclustering the data, as noted before (Shekhar et al., 2016). In *Step 2*, the subclusters within each cluster were assessed for differential expression, and iteratively merged as follows. We built a dendrogram on the subclusters (hierarchical agglomerative clustering with Euclidean distance, complete linkage), and tested the closest three pairs of leaves of the dendrogram for differential gene expression using the MAST test (Finak et al., 2015), corresponding to the three pairs of the most transcriptionally similar subclusters. If any pair did not show sufficient differential

expression (minimum 5 DE genes enriched both ways, fold-change > 1.4, p-value < 0.001, excluding mitochondrial transcripts and ribosomal protein genes, or “batch genes” defined above), the corresponding subclusters were merged. If any of the three pairs of subclusters were merged, we rebuilt the dendrogram using the new subclusters, and repeated this procedure until convergence. *Step 1* and *Step 2* resulted in the splitting of some of the original Louvain clusters. In *Step 3* we built a dendrogram on the new set of clusters and iteratively merged (as in *Step 2*, except with more stringent parameters) until convergence. During this final dendrogram-based merging round, we elected to merge transcriptionally proximal clusters using a more stringent rule. In addition to the criteria described above (5 DE genes enriched both ways, fold-change > 1.4, p-value < 0.001, excluding mitochondrial and ribosomal, or batch-specific genes), we required that the two clusters being tested be distinguished by ON/OFF expression (i.e. not just differences in expression levels) of at least one gene in each of them. More specifically, we required that each of the two clusters show at least one gene that was expressed in at least 4-fold higher proportion of cells the other cluster (e.g. 80% of cells in cluster 1 and 20% of cells in cluster 2, and 60% of cells in cluster 2 and 15% of cells in cluster 2).

7. Removing doublets and contaminants: Although dimensionality reduction and clustering were performed separately for each cell class after an initial segregation, we noted that the data was not entirely free of contaminants. For example, we found small PR-like or AC-like clusters when analyzing BCs in the fovea or periphery. These were likely doublets as these clusters also expressed BC-markers, and comprised <1-2% of the cells in the dataset. While such clusters appeared more BC-like in the initial stage, when all cell classes were analyzed together, their contaminant-like features became more apparent when analyzed only with pure BCs. These contaminant-like clusters were removed manually, and we verified that the final clusters for each cell-class expressed at most negligible levels of markers of other cell-classes.

Distinguishing M vs L Cones: The transcript sequences of *OPNIMW* and *OPNILW* are extremely similar across Old World Monkeys and humans (Nathans et al., 1986; Onishi et al., 2002). In the crab-eating macaque, there are only 26 single nucleotide polymorphism (SNPs) over 1,095 nucleotides of the full-length cDNA sequences of *OPNIMW* and *OPNILW*, located between exon 2 and exon 5 (Onishi et al., 2002). This made it difficult to unambiguously assign a short read (98bp) aligning to this locus to either *OPNIMW* and *OPNILW*, as SNPs could not be distinguished from sequencing errors. Thus, cells in the foveal cone cluster that expressed transcripts aligning to this locus, but not the S-opsin *OPNISW*, were initially tagged as “M/L cones.” However, these SNPs could be robustly distinguished from sequencing errors when aggregating across *all* the reads aligning to this locus (Figure S2A), and their positions were consistent with previous studies (Onishi et al., 2002). For each M/L cone cell, we then counted the *OPNIMW*- and *OPNILW*-specific SNPs (Figure S2B). ~70% of M/L cones in the data possessed reads that had at least one transcript covering one of the 26 SNPs. Cells with exclusively *OPNIMW*- and *OPNILW*-specific SNP counts were labeled M and L-cones, respectively, and the remainder were called “mixed” (Figures S2D,E).

Differential Expression Analysis: Differential expression (DE) tests for a gene between a pair of clusters or between a cluster and the remaining clusters were performed using the ‘MAST’ package (Finak et al., 2015), which implements a two-part ‘hurdle’ model to control for both technical quality and cross-animal variation. P-values were corrected for multiple hypotheses by controlling the false discovery rate (FDR) using the R function ‘p.adjust’.

We sought markers for a cluster by comparing it to other clusters only within the same class for DE genes, since class specific markers are well-defined (Table S1, Figs. 1C). To obtain highly specific markers for histological validation, we ranked significantly enriched (>1.5 -fold difference, FDR adjusted $p < 10^{-6}$) markers in increasing order of the proportion of background cells where they were detected (Table S3). Differential gene expression patterns were visualized either in dotplots (e.g. Figure 2B) or box-and-whisker plots (e.g. Figure 4B).

To identify fovea or periphery specific DE genes for cell types that had 1:1 matches across the two regions, we collected the corresponding cells from each region and applied MAST as before, while controlling for library size differences. In choosing candidates to pursue for validation, we also avoided weakly enriched genes that could be explained by large differences in abundance of a different cell class in the same region. For example, *RHO* and *GNAT1* were weakly enriched in peripheral RGCs, BCs and ACs, which likely reflects contamination from the predominant rods in the periphery.

Transcriptional Mapping of Foveal Types to Peripheral Types within A Class: We compared the consistency between foveal and peripheral clusters for HCs, BCs, ACs and RGCs within each cell class using a multi-class classification approach, as described before (Shekhar et al., 2016). For PRs and non-neuronal cells, the fovea vs periphery mapping was straightforward as the types could be easily identified based on known markers, and their transcriptional identities could be matched between the two regions.

For BCs, RGCs, ACs and HCs, we trained a multi-class classifier using Random Forests (RF) as well as the recently published Xgboost algorithm, which uses gradient boosted trees (Tianqi Chen, 2016) using the R packages ‘randomForest’ and ‘xgboost’, respectively. In each case, we trained the classifier on the peripheral cluster labels, using as features the common HVGs identified in the foveal and peripheral datasets as above (cluster-specific markers were not favored in any way). The training was performed on 50% of the cells in peripheral datasets and its performance was tested on the remaining, “held-out” dataset. (The foveal data was not used for training.) For each of the 4 classes (HCs, BCs, ACs and RGCs), the maximum error rate for types in the held-out data was $< 1\%$ (not shown), suggesting that the type-specific transcriptional signatures were robust, learnable, and not prone to overfitting.

The trained classifier was then used to assign each foveal cell a peripheral identity, in a manner that was agnostic to its foveal cluster identity. For each cell, we computed the “margin”, i.e the fraction of decision trees that vote for the “winner” type (majority vote). We excluded cells which fell into the following criterion,

$$\text{margin}(\text{cell}) < \min\left(0.6, \frac{5}{\# \text{ of types}}\right)$$

This was done to buffer against potential misclassifications due to a marginal majority. For all of the cell classes tested, < 4% of cells were excluded due to this criterion. The congruences of the foveal cluster identities and their peripheral matches were visualized as confusion matrices for each class (Figures 3A-D). A mapping between a foveal cluster F and a peripheral cluster P was considered 1:1 if > 85% of cells from F mapped to P , and if < 5% of cells from every other peripheral cluster mapped to F . Results from Xgboost and RF were extremely comparable, although the former was much more computationally efficient. The confusion matrices (Figures 3A-D) shown are equally-weighted aggregated results of Xgboost and RF.

To quantify the extent of 1:1 mapping between foveal and peripheral clusters we computed the Adjusted Rand Index (ARI), a measure of the similarity between two data clusterings, between the foveal cluster labels and the Xgboost/RF-assigned cluster labels using the R package “mclust”. The significance of the computed ARI values were evaluated by comparing against “null” ARI values computed by randomly permuting the cluster labels across cells. Mean and standard-deviation of null ARI values were computed by averaging 10,000 trials.

Comparison of Cluster Frequencies between the Fovea and the Periphery: We computed frequency differences of 1:1 matched types between the fovea and the periphery, and assessed their significance using a two sample t-test. For foveal types that were a mixture of multiple peripheral types (e.g. fRGC11), we assigned each foveal cell its matched peripheral cluster identity prior to comparison (see previous section). A fovea vs. periphery frequency comparison could not be meaningfully done for ACs, as CD90 non-uniformly labels ACs. We believe that this phenomenon likely explains the poor correlation between foveal and peripheral AC type frequencies ($r=0.12$); we could therefore not make a fair comparison of AC types between the regions. Interestingly, although the CD73- samples contributed only ~10% of all peripheral ACs, the AC type frequencies in these samples correlated much better with the foveal population ($r=0.79$; Figure S4B), and suggested that a few peripheral types (e.g. VIP+ ACs) were conspicuously underrepresented in the foveal samples (Figure 3C).

We quantified the compositional similarity between foveal and peripheral clusterings for each cell class by computing the Jensen-Shannon divergence (JSD), as follows. Let $P = \{p_1, p_2, \dots, p_n\}$ and $Q = \{q_1, q_2, \dots, q_n\}$ be the peripheral and foveal frequencies of n clusters of a cell class C , each of which sums to unity. Here n is the total number of clusters which is a sum of the 1:1 matched clusters, and the clusters unique to either region. Then the JSD between the fovea and the periphery for cell class C is computed as,

$$\text{JSD}(P | Q) = \frac{1}{2} D(P | M) + \frac{1}{2} D(Q | M)$$

Here, $M = \frac{1}{2}(P + Q)$, and $D(P | Q)$ is the Kullback-Leibler (KL) divergence between probability distributions P and Q as follows,

$$D(P | Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}$$

Note that while $D(P | Q) \neq D(Q | P)$, $JSD(P | Q)$ is symmetric in that $JSD(P | Q) = JSD(Q | P)$.

In comparing compositional similarity between foveal and peripheral cell classes, we used average frequencies across all replicates for foveal PRs, BCs, HCs, ACs and RGCs, but only BCs, HCs, ACs and RGCs in the periphery. This is because our collection procedure (Figure S1B) methods selected against peripheral PRs. We therefore used a published estimate of peripheral PR composition: 95% rods, 4.75% M/L cones and 0.25% S cones (Roorda et al., 2001; Wikler et al., 1990). In addition, we considered only the CD73⁻ samples for ACs because the CD90⁺ samples differentially excluded some AC types. For peripheral ACs, we only used their composition in the CD73⁻ samples (see above). We expect rods to be underrepresented in our peripheral samples due to our collection procedure.

Transcriptional Mapping of Cell Types across Species

Mouse: We transcriptionally related macaque cones, HC, BC, AC, and RGC types to their mouse counterparts from earlier publications using the same approach that related foveal to peripheral types, with minor modifications (see below). We conducted the following mappings:

1. 11 macaque peripheral BC clusters (BB/GB* was considered as a single cluster and then subclustered *post hoc*) were mapped to 15 mouse BC types published in our earlier study that identified these types molecularly and matched them to morphology (Shekhar et al., 2016)
2. 34 Macaque peripheral AC clusters were mapped to 21 mouse AC clusters identified in our earlier whole-retina study (Macosko et al., 2015), which represents an incomplete molecular characterization of ACs, estimated to contain >60 types based on morphological (Diamond, 2017) and molecular (JS, WY, KS, Ar, and Mallory Laboulaye, *unpublished*) diversity.
3. Macaque peripheral H1 and H2 were transcriptionally related to cells from a single cluster of mouse HCs in our previous whole retina study (Macosko et al., 2015) to determine whether mouse HCs were molecularly more similar to one of the two types.
4. 18 macaque peripheral RGC clusters were mapped to a recently published dataset of postnatal day 5 (P5) mouse RGCs, which identified 40 distinct molecular clusters (Rheaume et al., 2018).
5. As we did not sample peripheral S-cones, we related macaque foveal M/L and S cones to mouse cones from our previous full retina data (Macosko et al., 2015).

We computationally separated mouse cones into those that exclusively expressed *Opn1sw* or *Opn1mw* to enrich for pure S-cones and pure M-cones, respectively.

In all cases except HCs, we trained the classifier on the mouse types, and mapped the macaque cells on to these types. In each case 1-5 above, we trained RF/Xgboost classifiers on the training data (mouse for Cones, BCs, ACs, RGCs and macaque for HC) using as features the common set of highly variable genes (HVGs) in both the mouse and the macaque datasets among the 1:1 orthologues. Restricting the features to the common set of HVG orthologues was important for the performance of the classifier; in our initial tests, we found that using HVG orthologues identified only in one of the two species resulted in mappings that were more diffuse, likely due to genes that expressed in one species at high levels but, were completely absent in the other due to transcriptional rewiring. However, we emphasize that no manual curation of the feature set was performed to favor or disfavor certain genes based on their cell-type specific expression. We trained the classifier using 80% of the cells in the relevant mouse dataset, and validated it on the “held-out” cells in that dataset, which showed excellent performance in all cases. Next, we applied the classifier on the test set (macaque Cones, BCs, ACs, RGCs and mouse HCs), and assigned each cell an identity in the other species. As above, cluster identity of cells in the test set was not used to guide the mapping. Mappings of cell types across species were visualized using confusion matrices for Cones, ACs, BCs and RGCs (Figures 5A,C-E) and a barplot for HCs (Figure 5B). A mapping between a test cluster C_{test} and a training cluster C_{train} was considered 1:1 if $> 75\%$ of cells from C_{test} mapped to P , and if $< 5\%$ of cells from every other test cluster mapped to C_{train} .

Human: We related human BC clusters to macaque BC clusters using an approach similar to that described above for mouse. We trained a classifier on the macaque peripheral BC data, applied it to map each human BC to a macaque cluster, and visualized the resulting confusion matrix to identify 1:1 matches (Figure 6F).

Conservation of Transcription Factor (TF) Codes in Mouse and Macaque RGCs: We downloaded a curated list of human and mouse transcription factors from the TFCat Transcription Factor Database Catalog (www.tfcats.ca), CIS-BP database (<http://cisbp.cabr.utoronto.ca/>) and (Lambert et al., 2018), and used these to assemble a corresponding list of macaque transcription factors from our orthology list (see above). We filtered this list in two ways. **First**, we restricted the list to those TFs that had 1:1 mouse and macaque orthologs. **Second**, we only considered those TFs that were expressed (> 0 UMIs/cell) in $> 25\%$ of the cells in at least one cluster.

For each of the remaining TFs, we computed a TF specificity score (TFSS) among RGC types separately in macaque and mice. Specifically, for a transcription factor t , we computed a vector \vec{p}_t of the fraction of cells expressing t in every RGC cluster. For example, $\vec{p}_t = (0.7, 0, 0.3, \dots)$ implies that 70% of cells in RGC cluster 1, 0% of cells in cluster 2, 30% in cluster 3, and so on, express transcription factor t . We normalized \vec{p}_t to sum to 1, and computed the Shannon index of the resulting normalized vector,

$$TFSS(t) = - \sum_{clust} p_{t,i} \ln p_{t,i}$$

We computed $TFSS_{mouse}(t)$ and $TFSS_{macaque}(t)$ for all filtered transcription factors t ($n=93$), and compared them (Figure 5G). Note that constitutively expressed TFs like *Pou4f1/Six6* have low values on the plot, because of their low Shannon indices.

To compute TF co-expression scores and compare across species, we further filtered the 1:1 orthologous TF pairs to only include those TFs that are expressed in at least 15% of clusters in both species. For each species, we then binarized TF expression in RGC clusters as a matrix T , where $T(t, c) = 1$ if TF t is expressed in at least 25% of cells in cluster c . For every TF t , we computed its probability of expression in types, $\alpha(t)$, in any species as,

$$\alpha(t) = \frac{1}{N_c} \sum_c T(t, c)$$

Here, N_c refers to the number of clusters. The observed probability of co-expression of TFs t_1 and t_2 , $A(t_1, t_2)$, can be computed as,

$$A(t_1, t_2) = \frac{1}{N_c} \sum_c T(t_1, c) * T(t_2, c)$$

We compute the TF co-expression scores (TFCS) in either species as,

$$\begin{aligned} TFCS(t_1, t_2) &= \log\left(\frac{\text{Observed co-expression probability}}{\text{Expected co-expression probability}}\right) \\ &= \log\left(\frac{A(t_1, t_2)}{\alpha(t_1)\alpha(t_2)}\right) \end{aligned}$$

Figure S5G compares $TFCS_{mouse}(t_1, t_2)$ and $TFCS_{macaque}(t_1, t_2)$.

Evaluating Cell-Type Specific Expression of Disease Associated Loci: We assembled lists of genome-wide associated study genes for prominent retinal diseases that cause blindness – primary open angle glaucoma (POAG), primary angle closure glaucoma (PACG), cone-rod dystrophy (CRD), retinitis pigmentosa (RP), congenital stationary night blindness (CSNB), age-related macular degeneration (AMD) and diabetic macular edema (DME) from (Farrar et al., 2017; Fritsche et al., 2016; Graham et al., 2018; Wiggs and Pasquale, 2017; Zeitz et al., 2015).

To facilitate 1:1 comparison between the fovea and the periphery, we assigned each foveal cell a peripheral type identity using the RF/Xgboost mapping described above. We only considered those genes that were expressed in at least 20% of cells of at least one cell type, which resulted in 196 genes for further consideration. We note that this does not necessarily rule out the possibility of these genes turning on in specific cell types during disease

pathogenesis, just as the *expression* of a gene in a cell type at baseline does not necessarily imply vulnerability in disease.

We visualized expression patterns of the disease genes in our atlas using one of three visualizations:

1. Dotplots, which show cell-type resolution (as in Figure 7D),
2. Heatmaps showing relative expression of individual genes in major cell classes (Figures 7B, S7), and
3. Heatmap showing aggregated expression of groups of genes corresponding to individual diseases in major cell classes (Figure 7A). While 2. and 3. lack cell-type resolution, they do show statistical enrichment of genes and gene groups in specific cell classes, and differences in fovea and periphery, if they exist.

For each of the fovea and periphery, we summarized the expression of all disease genes (see above) across all cell types, by computing two matrices \mathbf{E} and \mathbf{F} . For gene g and cell-type c , $\mathbf{F}(g,c)$ was the fraction of cells in cell type c that had non-zero expression (> 0 UMIs) of g , and $\mathbf{E}(g,c)$ was the average number of transcripts in cells with non-zero expression. To remove outliers, we capped values in \mathbf{E} to the 99.5th quantile of expression. We then computed a matrix of gene expression scores for all genes across all cell types as the product of these two matrices,

$$\tilde{\mathbf{E}} = \mathbf{E}\mathbf{F}$$

For (2), we computed the expression strength $S(g,C)$ of each gene g in a cell class C (one of Rods, Cones, HCs, BCs, ACs, RGCs, and non-neuronal cells)

$$S(g,C) = \max_{c \in C} \tilde{\mathbf{E}}(g,c)$$

Here, we computed the maximum, rather than the average or aggregate, to highlight specific instances of highly cell-type specific expression of genes in a class. We combined $S_{fovea}(g,C)$ and $S_{periphery}(g,C)$, and visualized them as a single heatmap, where we z-scored the rows (corresponding to genes) to highlight relative expression strengths $\tilde{S}_{comb}(g,C)$. Specific rows are highlighted in **Figure 6I**. Here, the initial filtering of removing genes that were not expressed in at least 20% of cells in any cell type avoided spurious patterns.

For (3), we subset $\tilde{S}_{comb}(g,C)$ by disease and computed the mean relative expression strengths within each cell type group for each disease. This aggregate expression strength was visualized in Figure 7A.

In Figure S7, we visualized the relative expression strengths of genes grouped by disease as in Figure 7B, but after subdividing non-neuronal cells into MGs, Pericytes, Endothelial cells and Microglia to highlight differential expression among these types.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical methods and packages used for performing various types analysis in this work are cited where appropriate in the STAR Methods text. All the analysis was performing using the R statistical language in the RStudio IDE. Expression patterns of genes across cell clusters are shown in dotplots (e.g., Figure 2B), which simultaneously depict the fraction of cells in a cell cluster/type (column) that express a particular (row) based on the size of the dot, and the average number of transcripts in cluster based on the color scale, as indicated in the legends. In box and whisker plots depicting gene expression differences foveal and peripheral types, black horizontal line represents the median expression value, colored bars represent the interquartile range, and vertical lines span the minimum and maximum values. Stars indicate >2-fold changes based on mean expression values.

The concordance between two sets of clusters as assessed by the multi-class classification approach (see METHOD DETAILS) are visualized using a “Confusion Matrix” (e.g. foveal BCs vs peripheral BCs in Figure 3A, or macaque BCs vs mouse BCs in Figure 5E). The concordance between the two data clusterings is quantified using the Adjusted Rand Index (ARI).

Differential expression of genes between clusters were evaluated using the MAST test (Finak et al., 2015). We employed multiple hypothesis correction wherever significance was evaluated across multiple statistical tests, only considering genes more significant than False-discovery rate (FDR) threshold of 0.001. We provide FDR corrected p-values for specific instances of fovea vs. periphery differential expression reported in Figure 4.

DATA AND SOFTWARE AVAILABILITY

The following data files are available under the Gene Expression Omnibus (GEO) accession number GSE118480 – (1) Raw and processed data files for the macaque single-cell RNA-seq experiments, (2) Raw data files for the long-read bulk RNA-seq data from Tru-Seq libraries (for improving the transcriptomic reference, and (3) gtf file summarizing the transcriptome for *M. fascicularis* used in this study. Single cell data can be visualized in a single cell portal at https://portals.broadinstitute.org/single_cell. Instructions are provided in supplementary materials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank R. Born, A. Liu and J. Wiggs for advice; D. Juric for human tissue, and L. Gaffney, A. Hupalowska and E. Martersteck for assistance. This work was supported by grants from the NIH (MH105960, EY025840, EY028633, EY025555 and EY028625) and the BrightFocus Foundation (M2014055).

REFERENCES

Baden T, Berens P, Franke K, Roman Roson M, Bethge M, and Euler T (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature* 529, 345–350. [PubMed: 26735013]

- Bae JA, Mu S, Kim JS, Turner NL, Tartavull I, Kemnitz N, Jordan CS, Norton AD, Silversmith WM, Prentki R, et al. (2018). Digital Museum of Retinal Ganglion Cells with Dense Anatomy and Physiology. *Cell* 173, 1293–1306 e1219. [PubMed: 29775596]
- Blanks JC, and Johnson LV (1984). Specific binding of peanut lectin to a class of retinal photoreceptor cells. A species comparison. *Invest Ophthalmol Vis Sci* 25, 546–557. [PubMed: 6715128]
- Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, Jonas JB, Keeffe J, Leasher J, Naidoo K, et al. (2013). Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob Health* 1, e339–349. [PubMed: 25104599]
- Bringmann A, Syrbe S, Gorner K, Kacza J, Francke M, Wiedemann P, and Reichenbach A (2018). The primate fovea: Structure, function and development. *Prog Retin Eye Res*.
- Brooks MJ, Rajasimha HK, Roger JE, and Swaroop A (2011). Next-generation sequencing facilitates quantitative analysis of wild-type and *Nrl(-/-)* retinal transcriptomes. *Mol Vis* 17, 3034–3054. [PubMed: 22162623]
- Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36, 411–420. [PubMed: 29608179]
- Cajal SR (1893). La rétine des vertébrés. *Cellule* 9, 121–255.
- Chan TL, Martin PR, Clunas N, and Grunert U (2001). Bipolar cell diversity in the primate retina: morphologic and immunocytochemical analysis of a new world monkey, the marmoset *Callithrix jacchus*. *J Comp Neurol* 437, 219–239. [PubMed: 11494253]
- Cherry TJ, Wang S, Bormuth I, Schwab M, Olson J, and Cepko CL (2011). NeuroD factors regulate cell fate and neurite stratification in the developing retina. *J Neurosci* 31, 7365–7379. [PubMed: 21593321]
- da Silva S, and Cepko CL (2017). Fgf8 Expression and Degradation of Retinoic Acid Are Required for Patterning a High-Acuity Area in the Retina. *Dev Cell* 42, 68–81 e66. [PubMed: 28648799]
- Dacey DM (2004). Origins of perception: retinal ganglion cell diversity and the creation of parallel visual pathways In Gazzaniga MS (Ed.), *The Cognitive Neurosciences* (3rd ed., pp. 281–301). Cambridge: MIT Press.
- Diamond JS (2017). Inhibitory Interneurons in the Retina: Types, Circuitry, and Function. *Annu Rev Vis Sci* 3, 1–24. [PubMed: 28617659]
- Doan T, Azevedo AW, Hurley JB, Rieke F. (2009) Arrestin competition influences the kinetics and variability of the single-photon responses of mammalian rod photoreceptors. *J Neurosci*. 29:11867–79. [PubMed: 19776273]
- Euler T, Haverkamp S, Schubert T, and Baden T (2014). Retinal bipolar cells: elementary building blocks of vision. *Nat Rev Neurosci* 15, 507–519. [PubMed: 25158357]
- Farrar GJ, Carrigan M, Dockery A, Millington-Ward S, Palfi A, Chadderton N, Humphries M, Kiang AS, Kenna PF, and Humphries P (2017). Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum Mol Genet* 26, R2–R11. [PubMed: 28510639]
- Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 16, 278. [PubMed: 26653891]
- Fortunato S, and Barthelemy M (2007). Resolution limit in community detection. *Proc Natl Acad Sci U S A* 104, 36–41. [PubMed: 17190818]
- Fritsche LG, Igl W, Bailey JN, Grassmann F, Sengupta S, Bragg-Gresham JL, Burdon KP, Hebbbring SJ, Wen C, Gorski M, et al. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* 48, 134–143. [PubMed: 26691988]
- Gollisch T, and Meister M (2010). Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* 65, 150–164. [PubMed: 20152123]
- Graham PS, Kaidonis G, Abhary S, Gillies MC, Daniell M, Essex RW, Chang JH, Lake SR, Pal B, Jenkins AJ, et al. (2018). Genome-wide association studies for diabetic macular edema and proliferative diabetic retinopathy. *BMC Med Genet* 19, 71. [PubMed: 29739359]

- Greff R (1874). Die Mikroskopische Anatomie des Sehnerven und der Netzhaut. In Handbuch der gesamten Augenheilkunde 1.
- Grunert U, and Martin PR (1991). Rod bipolar cells in the macaque monkey retina: immunoreactivity and connectivity. *J Neurosci* 11, 2742–2758. [PubMed: 1715391]
- Grunert U, Martin PR, and Wässle H (1994). Immunocytochemical analysis of bipolar cells in the macaque monkey retina. *J Comp Neurol* 348, 607–627. [PubMed: 7530731]
- Haghverdi L, Lun ATL, Morgan MD, and Marioni JC (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 36, 421–427. [PubMed: 29608177]
- Haverkamp S, Haeseleer F, and Hendrickson A (2003). A comparison of immunocytochemical markers to identify bipolar cell types in human and monkey retina. *Vis Neurosci* 20, 589–600. [PubMed: 15088712]
- Haverkamp S, and Wässle H (2004). Characterization of an amacrine cell type of the mammalian retina immunoreactive for vesicular glutamate transporter 3. *J Comp Neurol* 468, 251–263. [PubMed: 14648683]
- Hendrickson A, Possin D, Vajzovic L, and Toth CA (2012). Histologic development of the human fovea from midgestation to maturity. *Am J Ophthalmol* 154, 767–778 e762. [PubMed: 22935600]
- Hendrickson A, Yan YH, Erickson A, Possin D, and Pow D (2007). Expression patterns of calretinin, calbindin and parvalbumin and their colocalization in neurons during development of Macaca monkey retina. *Exp Eye Res* 85, 587–601. [PubMed: 17845803]
- Hofer H, Carroll J, Neitz J, Neitz M, and Williams DR (2005). Organization of the human trichromatic cone mosaic. *J Neurosci* 25, 9669–9679. [PubMed: 16237171]
- Hoshino A, Ratnapriya R, Brooks MJ, Chaitankar V, Wilken MS, Zhang C, Starostik MR, Gieser L, La Torre A, Nishio M, et al. (2017). Molecular Anatomy of the Developing Human Retina. *Dev Cell* 43, 763–779 e764. [PubMed: 29233477]
- Johnson TV, and Martin KR (2008). Development and characterization of an adult retinal explant organotypic tissue culture system as an in vitro intraocular stem cell transplantation model. *Invest Ophthalmol Vis Sci* 49, 3503–3512. [PubMed: 18408186]
- Kay JN, Voinescu PE, Chu MW, and Sanes JR (2011). Neurod6 expression defines new retinal amacrine cell subtypes and regulates their fate. *Nat Neurosci* 14, 965–972. [PubMed: 21743471]
- Kent WJ (2002). BLAT--the BLAST-like alignment tool. *Genome Res* 12, 656–664. [PubMed: 11932250]
- Klump KE, Zhang AJ, Wu SM, and Marshak DW (2009). Parvalbumin-immunoreactive amacrine cells of macaque retina. *Vis Neurosci* 26, 287–296. [PubMed: 19435546]
- Kolb H, and Marshak D (2003). The midget pathways of the primate retina. *Doc Ophthalmol* 106, 67–81. [PubMed: 12675488]
- Kolesnikov AV, Rikimaru L, Hennig AK, Lukasiewicz PD, Fliesler SJ, Govardovskii VI, Kefalov VJ, and Kisselev OG (2011). G-protein betagamma-complex is crucial for efficient signal amplification in vision. *J Neurosci* 31, 8067–8077. [PubMed: 21632928]
- Krishnaswamy A, Yamagata M, Duan X, Hong YK, and Sanes JR (2015). Sidekick 2 directs formation of a retinal circuit that detects differential motion. *Nature* 524, 466–470. [PubMed: 26287463]
- Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, and Weirauch MT (2018). The Human Transcription Factors. *Cell* 172, 650–665. [PubMed: 29425488]
- Lammerding-Koppel M, Thier P, and Koehler W (1991). Morphology and mosaics of VIP-like immunoreactive neurons in the retina of the rhesus monkey. *J Comp Neurol* 312, 251–263. [PubMed: 1748731]
- Lassova L, Fina M, Sulaiman P, and Vardi N (2010). Immunocytochemical evidence that monkey rod bipolar cells use GABA. *Eur J Neurosci* 31, 685–696. [PubMed: 20384812]
- Lee BB, Shapley RM, Hawken MJ, and Sun H (2012). Spatial distributions of cone inputs to cells of the parvocellular pathway investigated with cone-isolating gratings. *J Opt Soc Am A Opt Image Sci Vis* 29, A223–232. [PubMed: 22330383]
- Liao HW, Ren X, Peterson BB, Marshak DW, Yau KW, Gamlin PD, and Dacey DM (2016). Melanopsin-expressing ganglion cells on macaque and human retinas form two morphologically distinct populations. *J Comp Neurol* 524, 2845–2872. [PubMed: 26972791]

- Linderman GC, Rachh M, Hoskins JG, Steinerberger S, & Kluger Y (2017). Efficient Algorithms for t-distributed Stochastic Neighborhood Embedding arXiv preprint arXiv: 1712.09005.
- Liu J, Reggiani JDS, Laboulaye MA, Pandey S, Chen B, Rubenstein JLR, Krishnaswamy A, and Sanes JR (2018). Tbr1 instructs laminar patterning of retinal ganglion cell dendrites. *Nat Neurosci* 21, 659–670. [PubMed: 29632360]
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas R, Kamitaki N, Martersteck EM, et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214. [PubMed: 26000488]
- Majumdar S, Wässle H, Jusuf PR, and Haverkamp S (2008). Mirror-symmetrical populations of wide-field amacrine cells of the macaque monkey retina. *J Comp Neurol* 508, 13–27. [PubMed: 18288700]
- Mao CA, Li H, Zhang Z, Kiyama T, Panda S, Hattar S, Ribelayga CP, Mills SL, and Wang SW (2014). T-box transcription regulator Tbr2 is essential for the formation and maintenance of Opn4/melanopsin-expressing intrinsically photosensitive retinal ganglion cells. *J Neurosci* 34, 13083–13095. [PubMed: 25253855]
- Mariani AP (1984). Bipolar cells in monkey retina selective for the cones likely to be blue-sensitive. *Nature* 308, 184–186. [PubMed: 6199677]
- Mariani AP, and Hokoc JN (1988). Two types of tyrosine hydroxylase-immunoreactive amacrine cell in the rhesus monkey retina. *J Comp Neurol* 276, 81–91. [PubMed: 2903868]
- Marioni JC, and Arendt D (2017). How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology. *Annu Rev Cell Dev Biol* 33, 537–553. [PubMed: 28813177]
- Masland RH (2012). The neuronal organization of the retina. *Neuron* 76, 266–280. [PubMed: 23083731]
- Masri RA, Percival KA, Koizumi A, Martin PR, and Grunert U (2016). Survey of retinal ganglion cell morphology in marmoset. *J Comp Neurol*. 10.1002/cne.24157.
- Meyer-Franke A, Kaplan MR, Pfrieger FW, and Barres BA (1995). Characterization of the signaling interactions that promote the survival and growth of developing retinal ganglion cells in culture. *Neuron* 15, 805–819. [PubMed: 7576630]
- Mills SL, and Massey SC (1999). AII amacrine cells limit scotopic acuity in central macaque retina: A confocal analysis of calretinin labeling. *J Comp Neurol* 411, 19–34. [PubMed: 10404105]
- Mitchell JF, Reynolds JH, and Miller CT (2014). Active vision in marmosets: a model system for visual neuroscience. *J Neurosci* 34, 1183–1194. [PubMed: 24453311]
- Nathans J, Thomas D, and Hogness DS (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232, 193–202. [PubMed: 2937147]
- Ng L, Hurley JB, Dierks B, Srinivas M, Salto C, Vennstrom B, Reh TA, and Forrest D (2001). A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nat Genet* 27, 94–98. [PubMed: 11138006]
- Onishi A, Koike S, Ida-Hosonuma M, Imai H, Shichida Y, Takenaka O, Hanazawa A, Komatsu H, Mikami A, Goto S, et al. (2002). Variations in long- and middle-wavelength-sensitive opsin gene loci in crab-eating monkeys. *Vision Res* 42, 281–292. [PubMed: 11809481]
- Pandey S, Shekhar K, Regev A, and Schier AF (2018). Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-Cell RNA-Seq. *Curr Biol* 28, 1052–1065 e1057. [PubMed: 29576475]
- Park SJ, Borghuis BG, Rahmani P, Zeng Q, Kim IJ, and Demb JB (2015). Function and Circuitry of VIP+ Interneurons in the Mouse Retina. *J Neurosci* 35, 10685–10700. [PubMed: 26224854]
- Peng YR, Tran NM, Krishnaswamy A, Kostadinov D, Martersteck EM, and Sanes JR (2017). Satb1 Regulates Contactin 5 to Pattern Dendrites of a Mammalian Retinal Ganglion Cell. *Neuron* 95, 869–883 e866. [PubMed: 28781169]
- Perry VH, and Silveira LC (1988). Functional lamination in the ganglion cell layer of the macaque's retina. *Neuroscience* 25, 217–223. [PubMed: 3393278]
- Pertea M, Kim D, Pertea GM, Leek JT, and Salzberg SL (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11, 1650–1667. [PubMed: 27560171]

- Provis JM, Dubis AM, Maddess T, and Carroll J (2013). Adaptation of the central retina for high acuity vision: cones, the fovea and the avascular zone. *Prog Retin Eye Res* 35, 63–81. [PubMed: 23500068]
- Rheume BA, Jereen A, Bolisetty M, Sajid MS, Yang Y, Renna K, Sun L, Robson P, and Trakhtenberg EF (2018). Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. *Nat Commun* 9, 2759. [PubMed: 30018341]
- Rodriguez AR, de Sevilla Muller LP, and Brecha NC (2014). The RNA binding protein RBPMS is a selective marker of ganglion cells in the mammalian retina. *J Comp Neurol* 522, 1411–1443. [PubMed: 24318667]
- Roorda A, Metha AB, Lennie P, and Williams DR (2001). Packing arrangement of the three cone classes in primate retina. *Vision Res* 41, 1291–1306. [PubMed: 11322974]
- Rosvall M, and Bergstrom CT (2008). Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A* 105, 1118–1123. [PubMed: 18216267]
- Rouso DL, Qiao M, Kagan RD, Yamagata M, Palmiter RD, and Sanes JR (2016). Two Pairs of ON and OFF Retinal Ganglion Cells Are Defined by Intersectional Patterns of Transcription Factor Expression. *Cell Rep* 15, 1930–1944. [PubMed: 27210758]
- Sanes JR, and Masland RH (2015). The types of retinal ganglion cells: current status and implications for neuronal classification. *Annu Rev Neurosci* 38, 221–246. [PubMed: 25897874]
- Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemes J, Goldman M, et al. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* 166, 1308–1323 e1330. [PubMed: 27565351]
- Sinha R, Hoon M, Baudin J, Okawa H, Wong RO, and Rieke F (2017). Cellular and Circuit Mechanisms Shaping the Perceptual Properties of the Primate Fovea. *Cell* 168, 413–426 e412. [PubMed: 28129540]
- Stella SL, Jr., Li S, Sabatini A, Vila A, and Brecha NC (2008). Comparison of the ontogeny of the vesicular glutamate transporter 3 (VGLUT3) with VGLUT1 and VGLUT2 in the rat retina. *Brain Res* 1215, 20–29. [PubMed: 18482716]
- Sweeney NT, James KN, Nistorica A, Lorig-Roach RM, and Feldheim DA (2017). Expression of transcription factors divides retinal ganglion cells into distinct classes. *J Comp Neurol*.
- Tianqi Chen CG (2016). XGBoost: A Scalable Tree Boosting System arXiv:160302754.
- Tsukamoto Y, and Omi N (2015). OFF bipolar cells in macaque retina: type-specific connectivity in the outer and inner synaptic layers. *Front Neuroanat* 9, 122. [PubMed: 26500507]
- Tsukamoto Y, and Omi N (2016). ON Bipolar Cells in Macaque Retina: Type-Specific Synaptic Connectivity with Special Reference to OFF Counterparts. *Front Neuroanat* 10, 104. [PubMed: 27833534]
- Vecino E, Rodriguez FD, Ruzafa N, Pereiro X, and Sharma SC (2016). Glia-neuron interactions in the mammalian retina. *Prog Retin Eye Res* 51, 1–40. [PubMed: 26113209]
- Wang Y, Smallwood PM, Cowan M, Blesh D, Lawler A, and Nathans J (1999). Mutually exclusive expression of human red and green visual pigment-reporter transgenes occurs at high frequency in murine cone photoreceptors. *Proc Natl Acad Sci U S A* 96, 5251–5256. [PubMed: 10220452]
- Wassle H, Dacey DM, Haun T, Haverkamp S, Grunert U, and Boycott BB (2000). The mosaic of horizontal cells in the macaque monkey retina: with a comment on bplexiform ganglion cells. *Vis Neurosci* 17, 591–608. [PubMed: 11016578]
- Wassle H, Grunert U, Martin PR, and Boycott BB (1994). Immunocytochemical characterization and spatial distribution of midget bipolar cells in the macaque monkey retina. *Vision Res* 34, 561–579. [PubMed: 8160377]
- Wiggs JL, and Pasquale LR (2017). Genetics of glaucoma. *Hum Mol Genet* 26, R21–R27. [PubMed: 28505344]
- Wikler KC, Williams RW, and Rakic P (1990). Photoreceptor mosaic: number and distribution of rods and cones in the rhesus monkey retina. *J Comp Neurol* 297, 499–508. [PubMed: 2384610]
- Wool LE, Crook JD, Troy JB, Packer OS, Zaidi Q, and Dacey DM (2018). Nonselective Wiring Accounts for Red-Green Opponency in Midget Ganglion Cells of the Primate Retina. *J Neurosci* 38, 1520–1540. [PubMed: 29305531]

- Yamada ES, Dmitrieva N, Keyser KT, Lindstrom JM, Hersh LB, and Marshak DW (2003). Synaptic connections of starburst amacrine cells and localization of acetylcholine receptors in primate retinas. *J Comp Neurol* 461, 76–90. [PubMed: 12722106]
- Zeitl C, Robson AG, and Audo I (2015). Congenital stationary night blindness: an analysis and update of genotype-phenotype correlations and pathogenic mechanisms. *Prog Retin Eye Res* 45, 58–110. [PubMed: 25307992]
- Zhang C, Yu WQ, Hoshino A, Huang J, Rieke F, Reh TA, and Wong ROL (2018). Development of ON and OFF cholinergic amacrine cells in the human fetal retina. *J Comp Neurol*.
- Zhang DQ, Zhou TR, and McMahon DG (2007). Functional heterogeneity of retinal dopaminergic neurons underlying their multiple roles in vision. *J Neurosci* 27, 692–699. [PubMed: 17234601]
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14049. [PubMed: 28091601]

HIGHLIGHTS

Macaque fovea and peripheral retina each contain >65 cell types

Most types correspond between regions but differ in proportions and gene expression

Greater conservation of interneuron than ganglion cell types between macaque and mice

Cell type- and region-specific expression of genes implicated in human blindness

Single cell based analysis provide a comprehensive molecular and cellular taxonomy of the primate retina

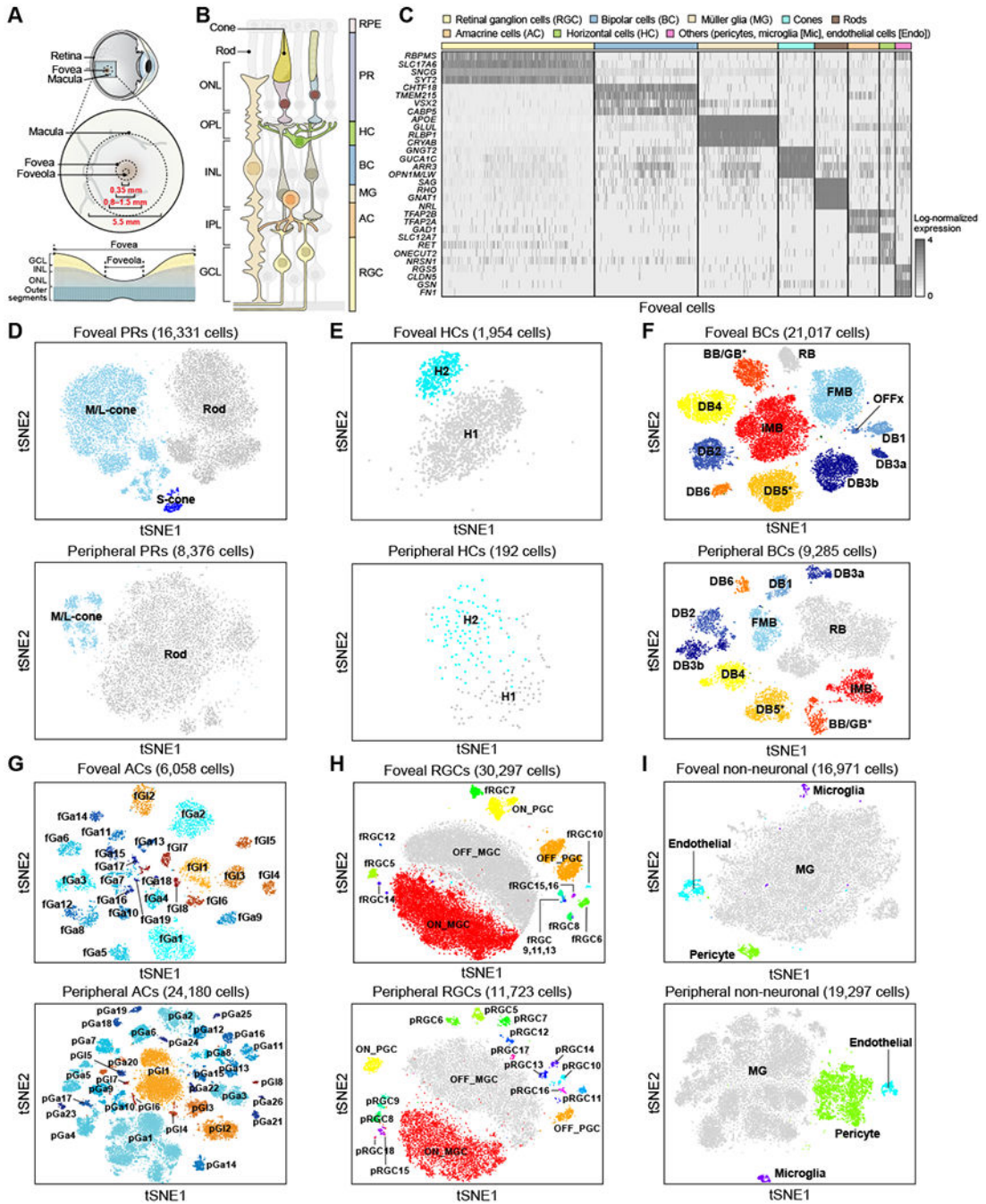


Figure 1: Single-cell profiling of peripheral and foveal cells from macaque retina
 A. (top) Sketch of a primate eye showing position of fovea and macula. (middle) Central region indicating diameters of the foveola (the foveal pit), fovea, and macula. (bottom) Sketch of a section through macaque fovea, showing foveal pit (red arrow) and displacement of ganglion cell layer (GCL) and inner nuclear layer (INL) cells.
 B. Sketch of peripheral retina showing its major cell classes – photoreceptors (PR), horizontal cells (HC), bipolar cells (BC), amacrine cells (AC), retinal ganglion cells (RGC)

and Muller glia (MG), outer and the inner plexiform (synaptic) layers (OPL, IPL), outer and inner nuclear layers (ONL, INL) and ganglion cell layer (GCL).

C. Expression patterns of class specific marker genes (rows) (Table S1) in single foveal cells (columns). Cells are grouped by their class (color bar, top). Plot shows randomly selected 10% of total cells. These signatures were used to separate peripheral cells into classes.

D-I. Visualization of foveal (top panel) and peripheral (bottom panel) PRs (D), HCs (E), BCs (F), ACs (G), RGCs (H), non-neuronal cells (I) using tSNE, a 2D non-linear transformation of high-dimensional data, that assigns proximal x-y coordinates to cells (points) with similar expression profiles. Individual cells are colored by their cluster assignments. Cluster labels, corresponding to *post hoc* assigned types, are indicated.

Although substructure is visible in D and I, this reflected batch effects and we were unable to detect subtypes by reanalysis. Because of their low density in the peripheral retina, only three S-cones were observed among the peripheral PRs, but these were not sufficient to form a separate cluster from M/L-cones. Foveal (f) and peripheral (p) AC clusters are divided into GABAergic (Ga) and Glycinergic (Gl) groups, and then numbered from largest to smallest cell number within each group. For RGCs, OFF and ON midget (MGC) and parasol (PGC) ganglion cells are labeled; clusters 5+ numbered from largest to smallest in the fovea (f) and periphery (p), respectively.

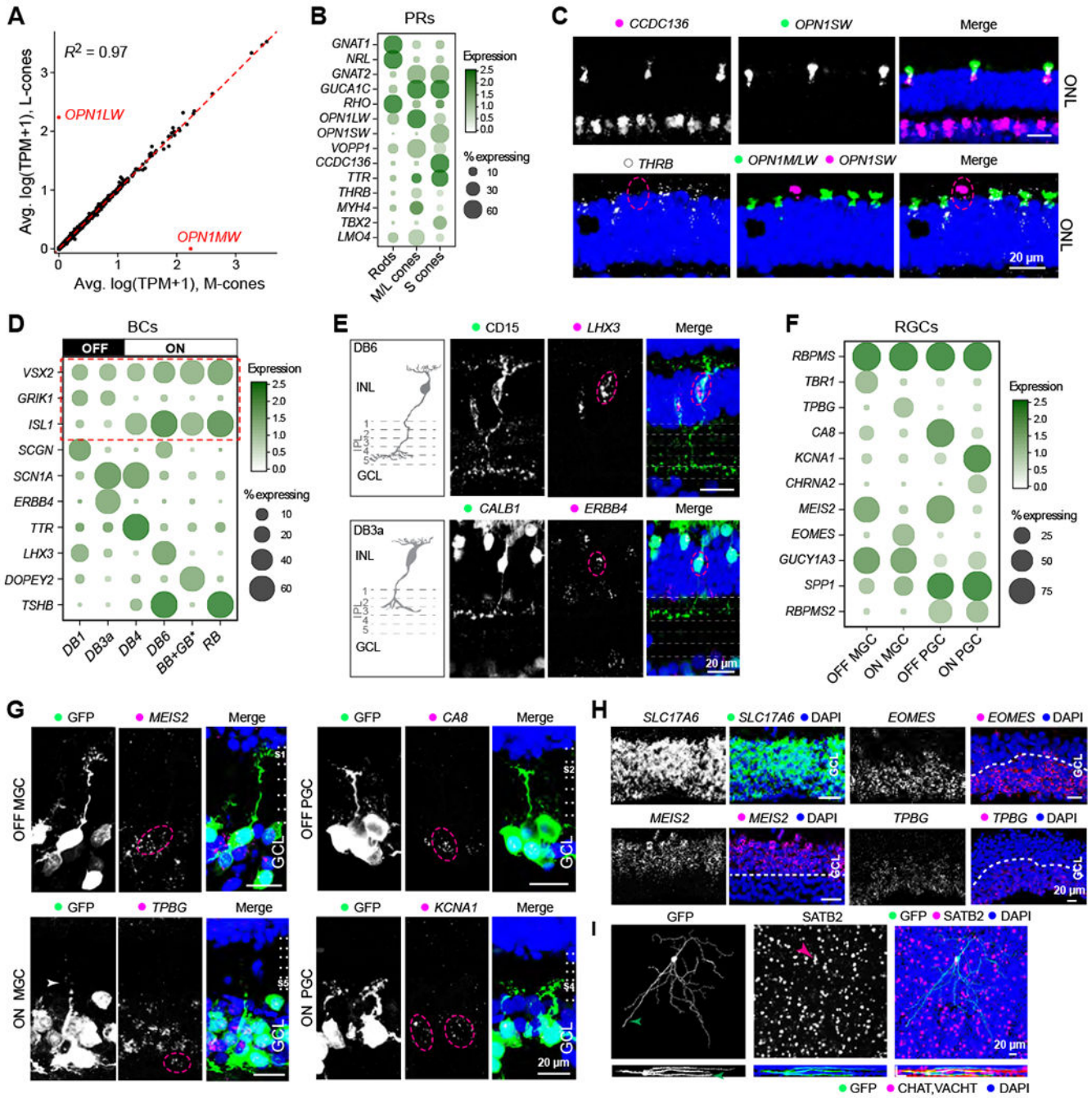


Figure 2: Matching scRNA-seq clusters to neuronal types of the primate retina

A. Comparison of average transcriptional profiles of foveal M-cones and L-cones. Each dot corresponds to a gene. No genes other than *OPN1MW* and *OPN1LW* differ significantly in expression levels (>1.2 fold at $p < 0.01$, MAST test) between the two cone types.

B. Dot plot showing expression of genes (rows) that distinguish photoreceptor (PR) types (columns) common to both the fovea and the periphery. The size of each circle is proportional to the percentage of cells expressing the marker (1 UMI), and its intensity depicts the average transcript count within expressing cells.

C. Validation of S cone specific gene *CCDC136* (upper) and M/L cone specific gene *THRB* (lower) by double and triple FISH with *OPN1SW* (S-opsin) and *OPN1MW/LW* (M/L-opsin) in the peripheral retina. Circle highlights an S cone.

D. Gene expression patterns of type-enriched markers for selected BC types. Red box highlights a pan-BC, a pan-ON-BC and a pan-OFF-BC marker. See Figure S2G and Table S2 for lists.

E. Validation of new markers for two BC types in peripheral retina by FISH combined with immunostaining. (top) DB6 (circle) cells, known to be CD15-positive, also express *LHX3*. (bottom) DB3a cell (circle) is *CALB1+ERBB4+*. IPL sublaminae (S1-S5) are demarcated by dashed lines. Sketches redrawn from (Tsukamoto and Omi, 2015, 2016).

F. Expression patterns of genes selectively enriched among foveal ON and OFF MGCs and PGCs.

G. Validation of markers for MGCs and PGCs from panel F, combining FISH with viral labeling (GFP) to show RGC morphology in the fovea.

H. Somata of ON MGCs (*TPBG+* and *EOMES+*) and OFF MGCs/PGCs (*MEIS2+*) are localized to the inner and outer halves (divided by dashed lines) of the ganglion cell layer, respectively, proximal to the fovea.

I. SATB2 positive SBC labeled by GFP-expressing virus in the peripheral retina.

Arrowheads indicate the axon (green) and soma. Bottom panels are rotations to show bistratified dendritic lamination.

Scale bar is 20 μm . DAPI staining is blue in C, E, G- I.

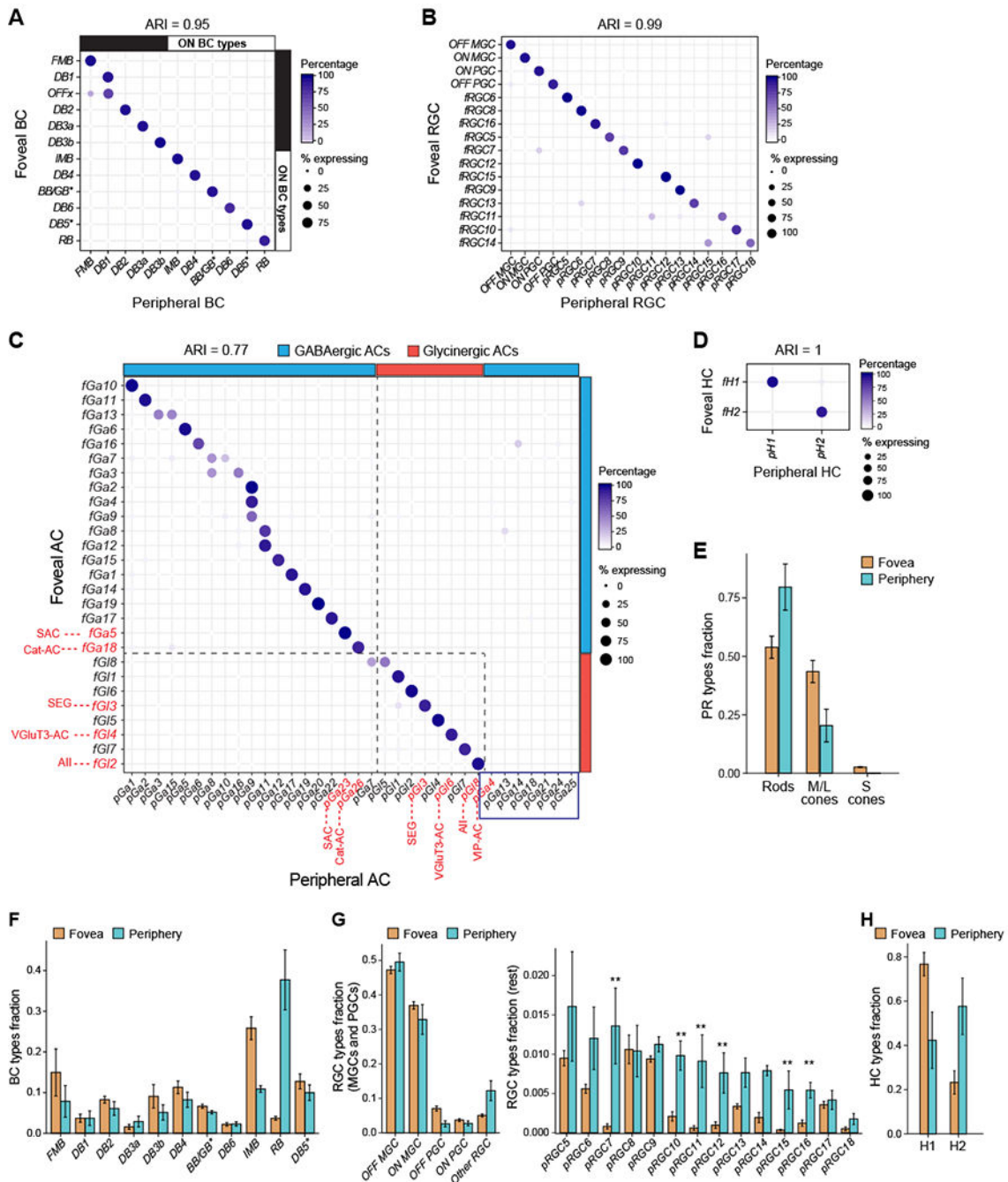


Figure 3: Correspondence between foveal and peripheral clusters

A-D. Transcriptional correspondence between foveal and peripheral clusters, summarized as “confusion matrices.” Circles and colors indicate the percentage of cells of a given foveal cluster (row) assigned to a corresponding peripheral cluster (column) by the classification algorithm trained on peripheral cells. A, BCs; B, RGCs; C, ACs; D, HCs. In panel A, bars on the top and right border mark ON and OFF BC subgroups. In C, bars mark GABAergic and Glycinergic AC subsets; key known types are labeled in red and types that might be periphery-specific are highlighted by blue box in C. Figure S4A provides molecular markers

for AC clusters. In panels A-D, the extent of 1:1 cluster matches are quantified by values of the Adjusted Rand Index (ARI), which range from close to 0 (random) to 1 (perfect 1:1 match). Empirical ARI values were highly significant for all classes, as compared to null ARI values (mean \pm SD) from random associations. BCs: $10^{-5} \pm 6 \times 10^{-4}$ RGCs: $-2 \times 10^{-5} \pm 6 \times 10^{-4}$ ACs: $-8.4 \times 10^{-6} \pm 3 \times 10^{-4}$, HCs: $3 \times 10^{-6} \pm 2 \times 10^{-4}$.

E-H. Comparison of cell type proportions between the fovea and the periphery (mean \pm SD, computed across biological replicates). E, PRs; F, BCs; G, RGCs; H, HCs. Foveal type OFFx and DB1 are grouped together as “DB1” due to their transcriptional similarity. To facilitate direct comparison of RGCs, each foveal type is assigned a peripheral identity from B. For ACs, see Figure S4A. Supervised analysis split fRGC11 and fRGC14 into two types each. RGC types underrepresented in the fovea are marked ** in G.

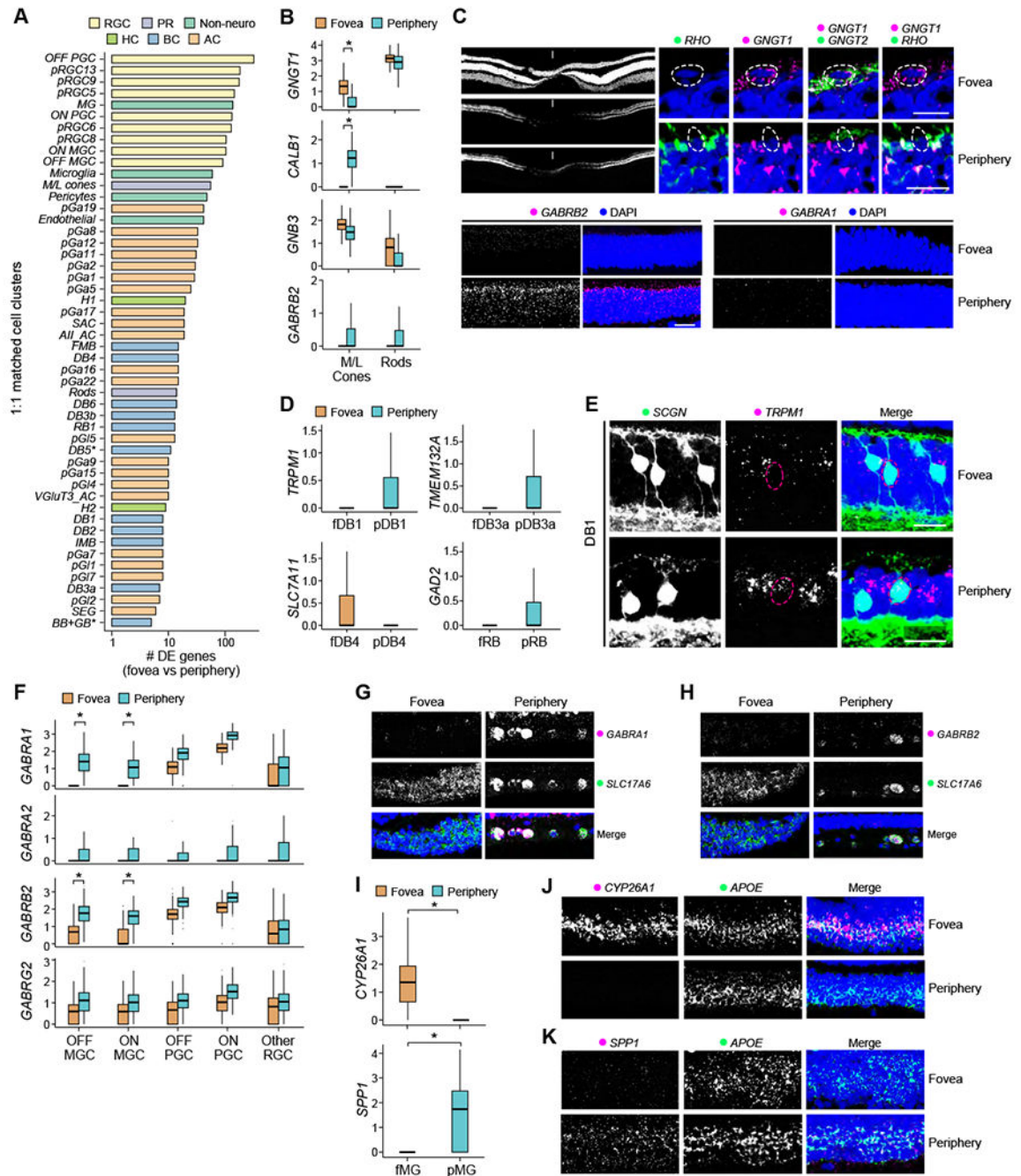


Figure 4: Differences in gene expression between foveal and peripheral cell types
 A. Barplot showing the number of DE genes per matched cluster between fovea and periphery (log-fold change >2, $p < 10^{-5}$, MAST test). Bars are labeled based on the corresponding peripheral cluster (except known types), colored by cell class and arranged in decreasing order. Only clusters with 50 cells in both the fovea and periphery are shown.
 B. Box and whisker plots show examples of gene expression differences foveal and peripheral PRs. p -values were calculated using the MAST package. Black horizontal line, median; Bars, interquartile range; vertical lines, minimum and maximum. All differences

between fovea and periphery in B,D,F,J are statistically significant with $p < 1e-19$. Stars indicate >2 -fold changes based on means.

C. In situ validation for B. (upper) Foveal but not peripheral cones express *GNPT1*. Arrow indicates foveal center. Circles highlight cones; asterisks highlight rods (*RHO+*). (lower) Peripheral PRs express higher *GABRB2* than foveal PRs; neither type expresses *GABRA1*.

D. Same as B, for BCs.

E. In situ validation for D. *TRPM1* is expressed in peripheral but not foveal DB1. Both are SCGN+.

F. Same as B, for RGCs.

G., H. In situ validation for F. *GABRA1* (G) and *GABRB2* (H) expression in foveal and peripheral RGCs (*SLC17A6+*).

I. Same as B, for MGs.

J., K. In situ validation for I. *CYP26A1* (J) and *SPP1* (K) are selectively expressed by foveal and peripheral MGs respectively. *APOE* is a pan-MG marker.

Scale bars, 20 and 300 μ m as indicated. DAPI staining is blue in C,E,G,J.

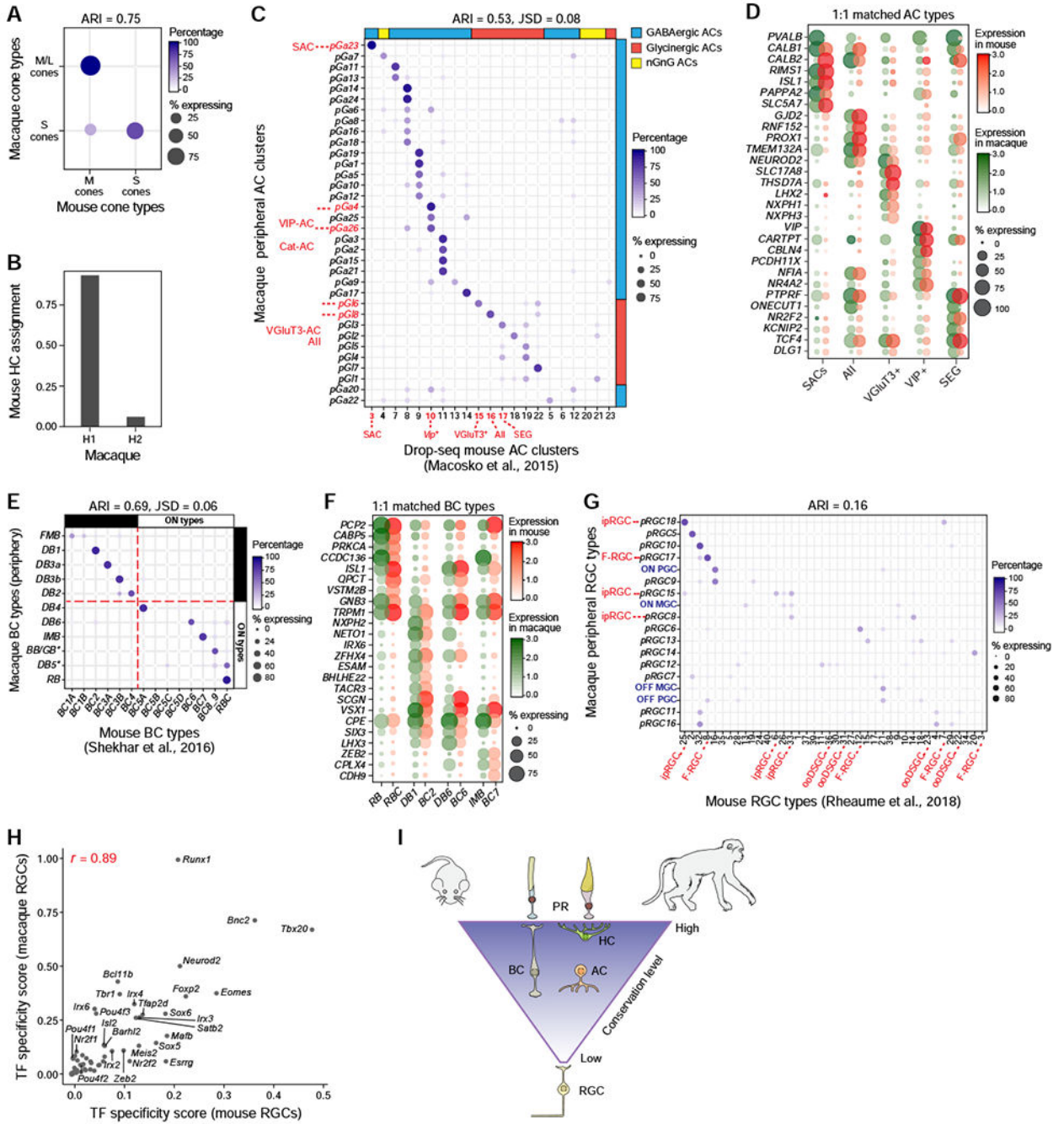


Figure 5: Conservation of retinal cell types between mouse and macaque

A. Transcriptional correspondence between macaque (rows) and mouse cone types (from Macosko et al., 2015) (columns). Only mouse cones expressing M- or S-opsin but not both were used for this comparison.

B. Supervised classification shows that HCs are more closely related to macaque H1 than to H2 cells.

C. Transcriptional correspondence between macaque peripheral AC clusters and mouse AC clusters (from Macosko et al., 2015). 1:1 mapping of multiple macaque AC clusters reflects

the incomplete resolution of AC types in published mouse data. Known types that map 1:1 are indicated (red).

D. Example of orthologous gene expression patterns in matched macaque-mouse AC types

E. Transcriptional correspondence between macaque peripheral BC types and mouse BC types (from Shekhar et al., 2016). 9 out of 11 macaque BC types map preferentially to a single mouse type. Crossed red lines highlight correspondence of OFF and ON groups.

F. Example of orthologous gene expression patterns in matched macaque-mouse BC types.

G. Transcriptional correspondence between macaque peripheral RGC clusters and mouse RGC clusters (from Rheaume et al., 2018). Select mouse and macaque types are indicated.

H. Despite poor 1:1 correspondence of RGC types across the two species (G), transcription factors (TFs) that exhibit restricted expression among subgroups are similar, based on high Pearson correlation coefficient (0.89) of their specificity scores (STAR Methods) between macaque and mouse.

I. Schematic showing decreasing molecular conservation of types within cell classes from outer to inner retina.

For panels A, C, E, G null ARI values (mean \pm SD) for a random model are PRs:

$1.7 \times 10^{-4} \pm 9 \times 10^{-3}$ ACs: $2 \times 10^{-5} \pm 10^{-3}$ BCs: $3 \times 10^{-4} \pm 9 \times 10^{-4}$ RGCs: $2 \times 10^{-6} \pm 7 \times 10^{-5}$ Cross-species compositional similarity for BCs and ACs was quantified using the JSD metric. To avoid enrichment biases (Figure S4B), only macaque ACs from CD73- samples were considered.

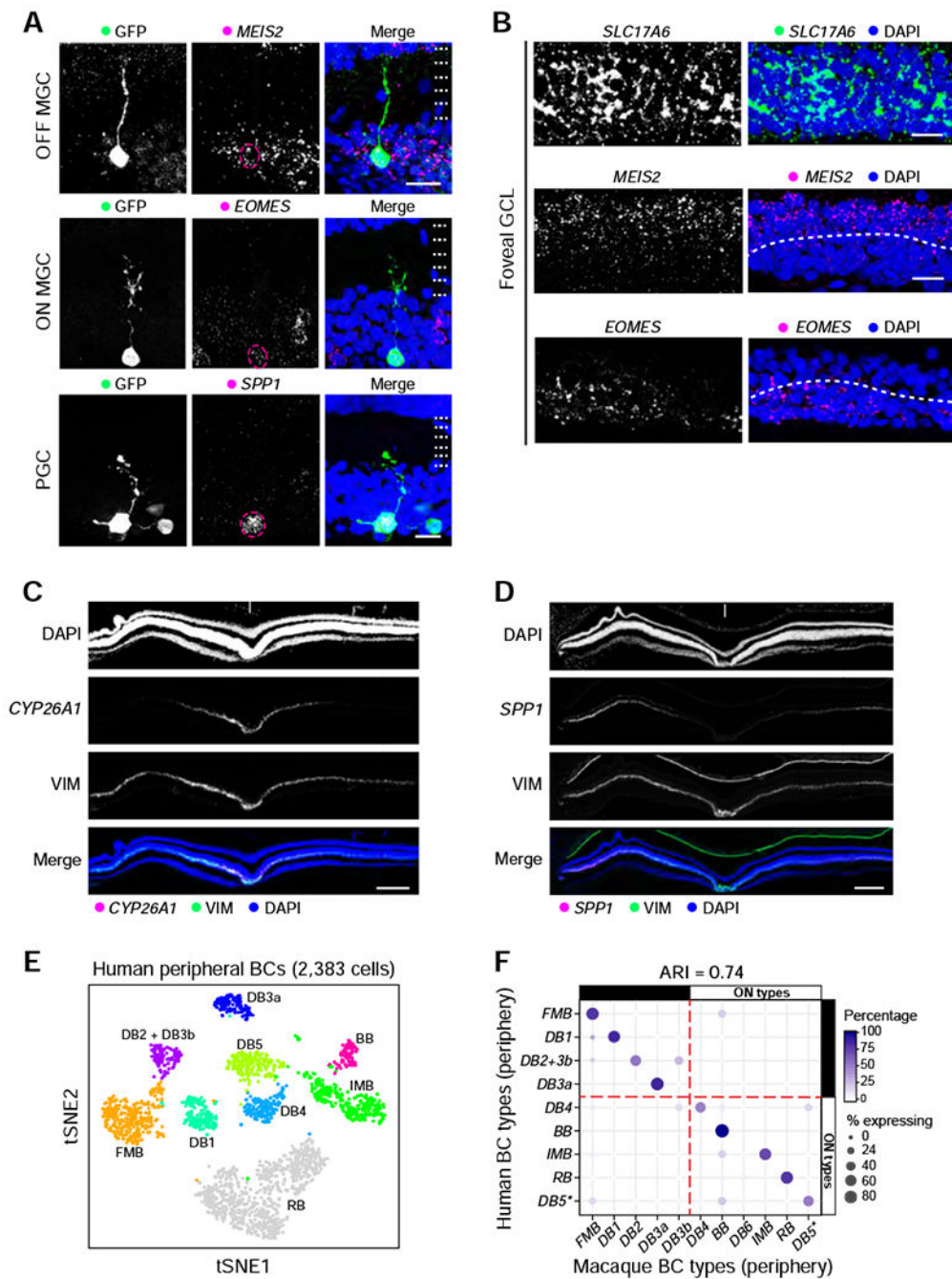


Figure 6: Conservation in marmosets and humans

A. Biolistic labeling combined with FISH shows *MEIS2*+ OFF-MGs (top), *EOMES*+ ON-MGs (middle), and *SPP1*+ PGCs (bottom) in the marmoset fovea.

B. The exclusive expression of *MEIS2* and *EOMES* in the marmoset foveal GCL layer.

C,D. *CYP26A1* (C) and *SPP1* (D) are selectively expressed by foveal and peripheral Müller glia respectively in marmoset retinas. Arrows indicate the center of the fovea.

E. t-SNE visualization of human peripheral BCs. Representation as in Figure 1D-I. Clusters are labeled based on their correspondence to macaque BC types (panel F).

F. Transcriptional correspondence between human (rows) and macaque (columns) peripheral BC types. Representation as in Figure 3A. Each human BC cluster is labeled here and in panel E retrospectively based on its most similar macaque BC.

Arrows indicate the center of fovea. Scale bars, 20 μ m and 300 μ m as indicated. DAPI staining is blue in A-D.

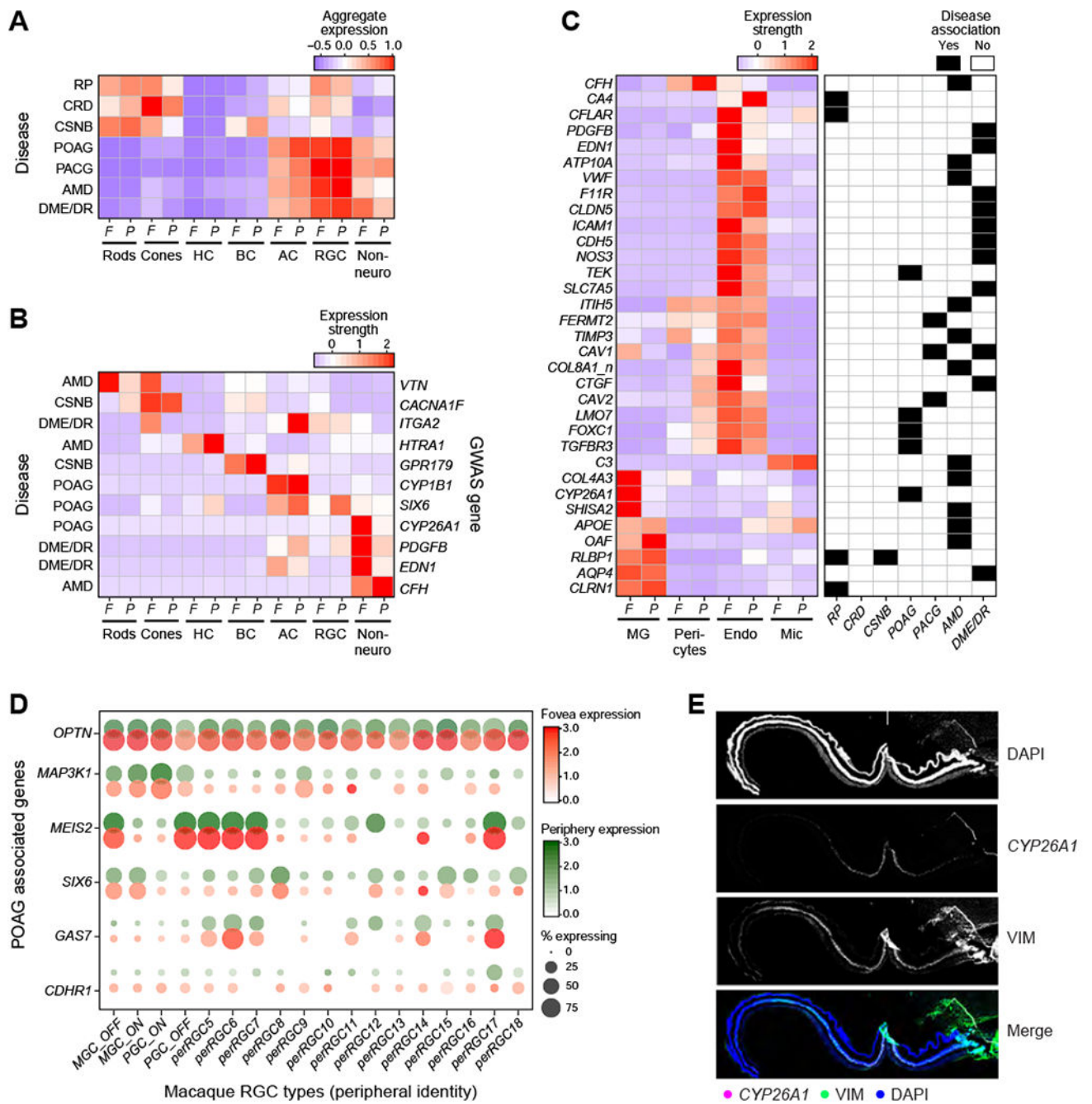


Figure 7: Cell-type and region-specific expression patterns of human retinal disease associated genes

A. Aggregated expression of disease-associated genes in foveal and peripheral cell classes. Disease groups are RP – retinitis pigmentosa; CRD – cone-rod dystrophy; CSNB – congenital stationary night blindness; POAG – primary open angle glaucoma; PACG – primary angle closure glaucoma; AMD – age-related macular degeneration, and DME/DR – diabetic macular edema and retinopathy.

B. Expression patterns of a subset of retinal-disease associated genes by cell class (columns), as in panel A. The primary disease associated with each gene is indicated on the right. See Figure S7 for full list.

C. Expression patterns of specific retinal-disease associated genes (rows) by non-neuronal types in the fovea and the periphery (columns).

D. Expression of glaucoma associated genes in RGC types in the fovea and the periphery.

E. The POAG susceptible gene, *CYP26A1*, shows its specific expression in human foveal müller glia.

Arrows indicate the center of fovea. Scale bars, 300 μm .