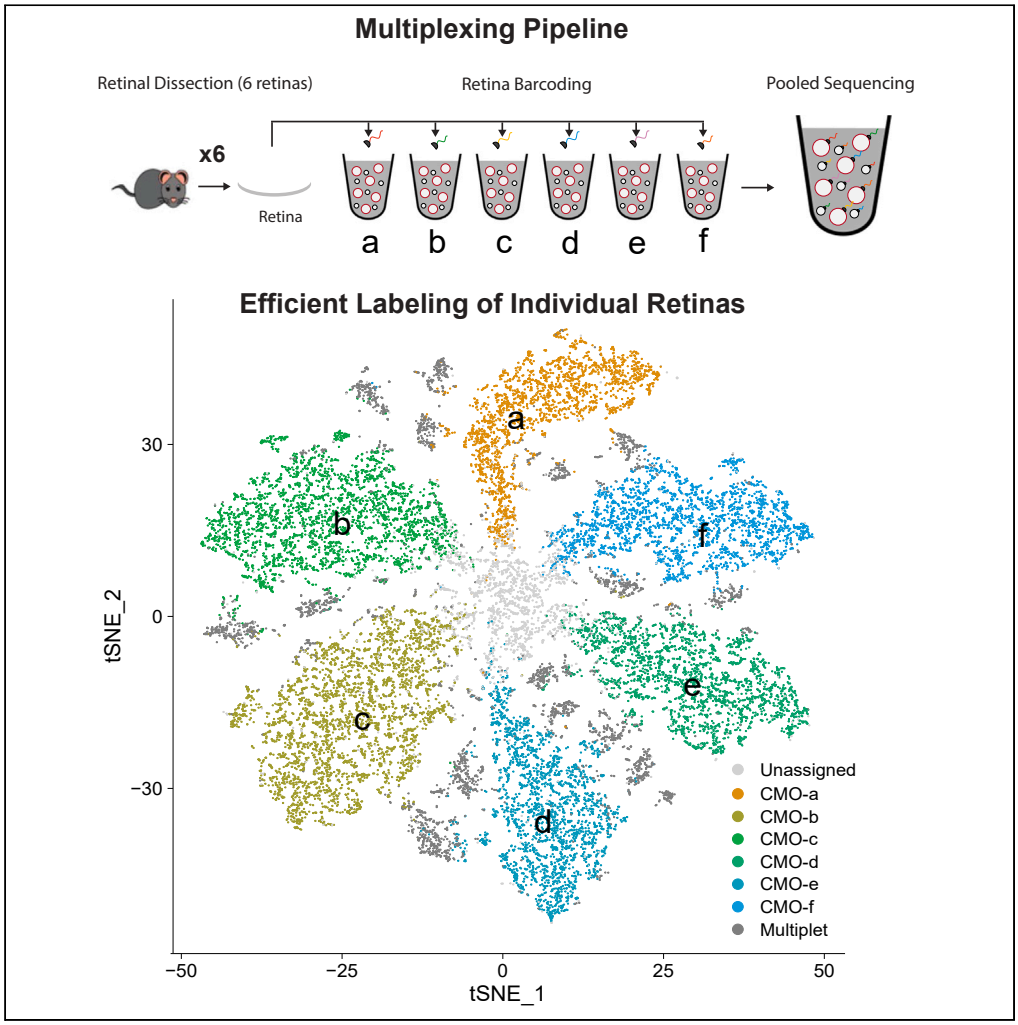
iScience

CellPress
OPEN ACCESS

Article

Sample multiplexing for retinal single-cell RNA sequencing



Justin Ma, Ting-Kuan Chu, Maria Polo-Prieto, ..., Graeme Mardon, Benjamin J. Frankfort, Nicholas M. Tran

nicholas.tran@bcm.edu

Highlights

Transcriptomic profiling of rare neuronal populations in individual retinas

Detection of cell multiplets was improved by sample barcoding

RGC types showed consistent distribution and gene expression across retinas

Refinement of RGC atlas resolved two pairs of transcriptionally similar types

Ma et al., iScience 27, 111250
November 15, 2024
https://doi.org/10.1016/j.isci.2024.111250

iScience

CellPress
OPEN ACCESS

Article

Sample multiplexing for retinal single-cell RNA sequencing

Justin Ma,[1] Ting-Kuan Chu,[2] Maria Polo-Prieto,[3] Yong H. Park,[4] Yumei Li,[2] Rui Chen,[2] Graeme Mardon,[1,2,3,4] Benjamin J. Frankfort,[3,4] and Nicholas M. Tran[2,5,*]

SUMMARY

Rare cell populations can be challenging to characterize using microfluidic single-cell RNA sequencing (scRNA-seq) platforms. Typically, the population of interest must be enriched and pooled from multiple biological specimens for efficient collection. However, these practices preclude the resolution of sample origin together with phenotypic data and are problematic in experiments in which biological or technical variation is expected to be high (e.g., disease models, genetic perturbation screens, or human samples). One solution is sample multiplexing whereby each sample is tagged with a unique sequence barcode that is resolved bioinformatically. We have established a scRNA-seq sample multiplexing pipeline for mouse retinal ganglion cells using cholesterol-modified oligos. We utilized the enhanced precision of this dataset to investigate cell type distribution and transcriptomic variance across retinal samples. Additionally, we demonstrate that our multiplexed dataset can be useful for the identification of multiplets in non-labeled samples, a common challenge in scRNA-seq analysis.

INTRODUCTION

Single-cell RNA sequencing (scRNA-seq) is a powerful tool for studying cell type- or state-specific transcription in healthy and diseased tissues.[1–3] Widely used microfluidic droplet-based scRNA-seq platforms enable profiling of transcriptomes from thousands of cells per experiment.[4,5] However, when working with rare or low abundance cell populations, it may not be feasible to obtain the required input cell number for these platforms to run efficiently. Cells can be pooled from multiple samples for each experiment, but this precludes analysis of gene expression differences across individual samples. This limitation is particularly challenging for experiments where gene expression is expected to differ across biological replicates (e.g., disease models with variable phenotypes, high-throughput genetic perturbation screens, or human patient samples). A solution for tracking sample identity in pooled scRNA-seq collections is sample multiplexing, whereby cells from each sample are tagged with unique next-generation sequencing (NGS) barcodes prior to pooling.[6–10] Sample barcodes are sequenced in parallel to the transcriptome, and sample identities of each cell are extracted bioinformatically. Here, we demonstrate the utility of using sample multiplexing to analyze gene expression patterns of rare cell types.

To establish our sample multiplexing pipeline, we focused on retinal ganglion cells (RGCs), which are the projection neurons that connect the retina to the rest of the brain. We previously generated a cellular atlas of mouse RGC types by scRNA-seq, identifying 46 types.[11–13] Since RGCs account for <1% of all retinal cells, our previous experiments pooled RGCs from ~6 to 8 retinas for each scRNA-seq collection, preventing the assessment of sample-specific gene expression.[12,14–16] In this study, we tagged RGCs from individual retinas using cholesterol-modified oligos (CMOs), which utilize a barcode attached to a lipid scaffold that can conjugate to cell surfaces.[7] CMO labeling efficiently tagged retinal cells and readily integrated with our RGC purification methods. CMO labeling did not impact transcriptome quality and enabled evaluation of key biological features such as sex-specific differences and assessment of cell type frequencies in individual retinas. Our analytical pipeline also highlights important considerations for sample demultiplexing, determination of labeling efficiency, identification of cell multiplets, among other features. Collectively, our studies demonstrate that CMO labeling enabled the resolution of sample origin in scRNA-seq data. Our research framework optimizes scRNA-seq approaches for studying RGCs in disease models and other experimental conditions and is adaptable to study rare cell populations in other tissues. Further, our CMO-labeled dataset enabled enhanced identification of multiplets in a non-labeled sample, demonstrating a generalizable utility. As scRNA-seq is increasingly being used to evaluate transcriptomic shifts in disease and development, sample multiplexing represents a useful tool that can enhance the precision and interpretability of analysis.

[1]Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
[3]Department of Ophthalmology, Baylor College of Medicine, Houston, TX, USA
[4]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
[5]Lead contact
*Correspondence: nicholas.tran@bcm.edu
https://doi.org/10.1016/j.isci.2024.111250

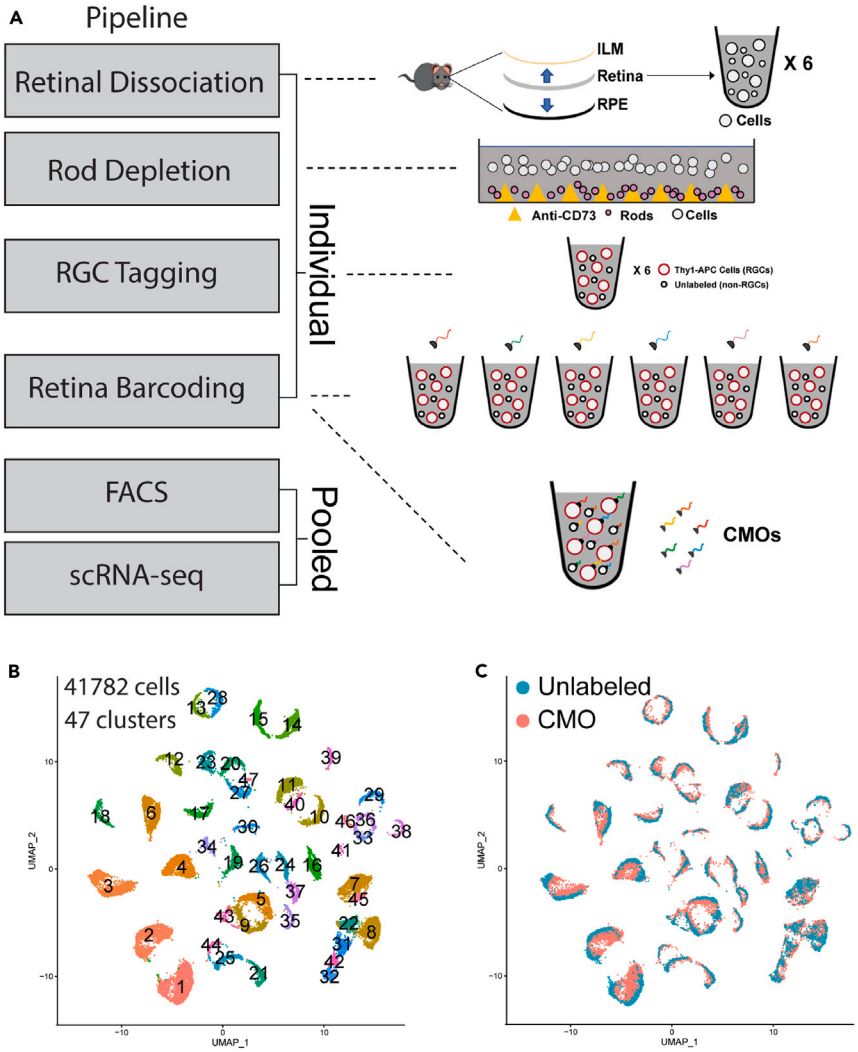


Figure 1. CMO labeling enables tracking of retinal sample origin in scRNA-seq

(A) Retinas from *Vglut2-Cre;Ai9* mice were enzymatically dissociated and processed in parallel. Rods were depleted by CD73 negative-immunopanning, and RGCs were labeled with an anti-CD90.2/Thy1.2 antibody conjugated to APC. Each retina was labeled with a unique CMO barcode and then pooled for RGC enrichment by FACS (TdTomato+; C90.2/Thy1.2+). Collected cells were processed for scRNA-seq.

(B) UMAP showing clustering of processed scRNA-seq dataset (6 collections) yielded 41,782 high-quality RGC transcriptomes grouped into 47 clusters. This included five CMO-labeled experiments (27 retinas) and one "Unlabeled" experiment (8,706 cells), which was not labeled with CMOs.

(C) All clusters contained both CMO-labeled and "Unlabeled" cells, demonstrating transcriptional similarity. ILM, inner limiting membrane; RPE, retinal pigment epithelium (see also Figures S1, S2 and Table S1).

RESULTS

CMO labeling of mouse retinal cells did not impact scRNA-seq quality

To enable tracking of sample identity of RGCs in scRNA-seq collections, we integrated CMO labeling into our RGC purification protocols[12,13] (Figure 1A). For each collection, up to six individual retinas from adult *Vglut2-cre;Ai9* mice were enzymatically dissociated and processed in parallel. We modified our RGC purification protocol to incorporate a negative anti-CD73 immunopanning step to deplete rod photoreceptors.[17] Rod-depletion removed ~60% of the total cells from each sample and reduced the amount of time needed for subsequent fluorescence-activated cell sorting (FACS) steps (Figures S1A and S1B). Each sample was incubated with a unique CMO barcoding tag and pooled prior to FACS. RGCs (CD90.2+; TdTomato+) were purified by FACS for scRNA-seq (10× Genomics) (Figure S1D). Consistent with previous experiments, retinal cells maintained high viability (~77% ± 8.3%) throughout collections and sequenced cells were ~83% ± 9.4% RGCs (Figure S1C and Table S1). As an additional control, we performed a non-CMO-labeled, rod-depleted pooled RGC collection (Exp0_Unlabeled), which similarly maintained high viability and RGC purity (Table S1).

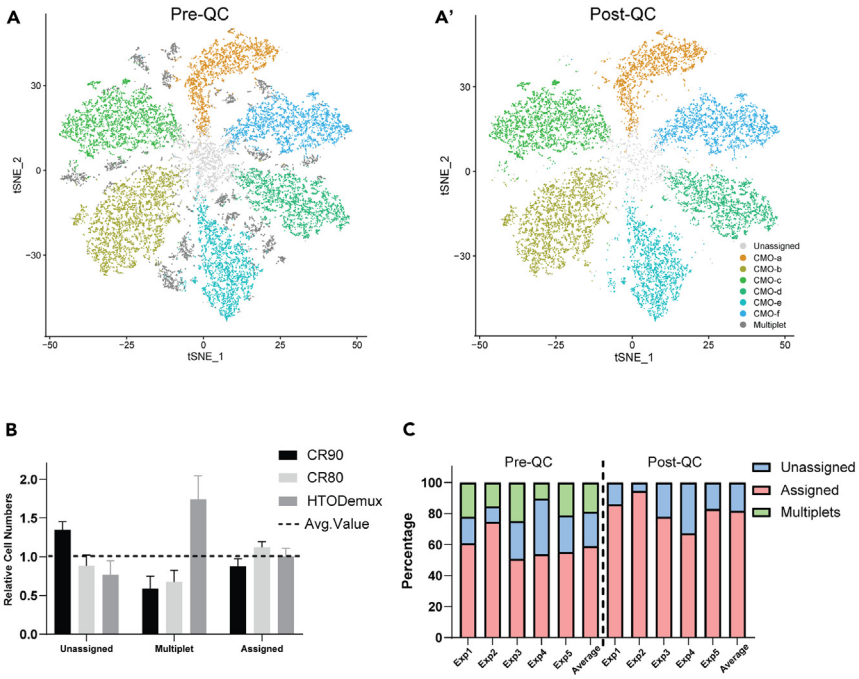


Figure 2. Demultiplexed retinas retain a high percentage of high-quality cells

(A) t-distributed stochastic neighbor embedding (tSNE) showing clustering of cells from experiment 2 (6 retinas) based on CMO-barcode reads. (A') Multiplets and most "Unassigned" cells were removed by quality control (QC) filters. Resulting clusters largely represent cells labeled with one CMO-barcode, which are predicted to be derived from the same retinal sample.

(B) Comparison of the HTODemux assignment method to the Cell Ranger (CR) multiplex pipeline (CR90 and CR80 represent 90% and 80% confidence interval limits used, respectively). Assignment percentages were calculated for each experiment ($n = 5$) and normalized to the average value for each of the 3 categories. Data are represented as normalized mean +/− standard deviation (SD).

(C) Comparison of assignment percentages pre- and post-QC filtering. For each side, columns 1–5 depict the percentage of cells represented by one of the three assignments for each experiment. On average, 81.6% ± 9.1% of the cells that passed QC and multiplet filters were assigned. All experiments are the average of their 6 retinas, with the exception of Exp4, which has 3 retinas (see also Figure S3 and Table S2).

Overall, we generated 41,782 high-quality RGC transcriptomes by scRNA-seq following quality control (QC) filtering, removal of non-RGCs, and removal of multiplets (Figure 1B). This included 33,076 RGCs collected from 27 CMO-labeled retinas (Table S2). Excluding Exp0 and 1, which had a lower cell input for scRNA-seq, we obtained a mean recovery of 1,373 RGCs per retina, ranging from ~1,006 to ~2,331 per retina (Figure S1E). Unsupervised clustering identified 47 transcriptionally related groups. Clustered data recapitulated RGC cell-type-specific gene expression patterns previously defined in the reference RGC atlas (Figure S2).[12] CMO labeling did not impact clustering as non-CMO-labeled cells co-clustered with CMO-labeled cells (Figure 1C). We detected 10 genes that were significant differentially expressed genes (DEGs) between CMO-labeled and Unlabeled cells (Figure S1F). However, it is unclear whether these genes have biological significance or are subtle batch effects, as most of these genes are related to the mitochondria or histones and were inherently highly variable between retinas regardless of preparation method (Table S6). Thus, we concluded that incorporation of anti-CD73 immunopanning and CMO labeling steps did not interfere with RGC purification and impact scRNA-seq data quality, or gene expression.

Tracking retinal sample origin with CMO labeling

The sample identity of CMO-labeling retinal cells was determined using computational methods. For each collection, cells were clustered based on detection of their unique CMO barcodes (Figures 2A and S3). We tested two methods to assign sample identity to cells: Cell Ranger Multiplexing Pipeline and the hashtag oligos Demultiplexing package (HTODemux, Seurat)[6,18,19] (Figure 2B). Both methods distinguish CMO tags from background noise by fitting to a binomial distribution, with the user setting a confidence threshold. We tested multiple thresholds for each method, classifying cells as "Assigned," Multiplets," or "Unassigned," based on the detection of one, multiple, or no CMO barcodes, respectively (Figures 2A–2C and Table S2). Assignments largely overlapped between methods; however, HTODemux identified a greater number of multiplets (Figure 2B). For subsequent analyses, we utilized the more stringent HTOdemux assignments. Assignment rate on average was ~82%, ranging from 67% to 94% across experiments (Figures 2C and Table S2).

As a secondary validation of sample identity assignments, we assessed the expression patterns of sex-related genes in experiments that included CMO-labeled retinas from male and female mice. Expression of genes located on the X (*Xist*) and Y (*Ddx3y, Eifs3y*) chromosomes was enriched specifically in cells assigned to samples of the appropriate sex (Figure 3). Multiple other sex-linked genes including *Gm21887*

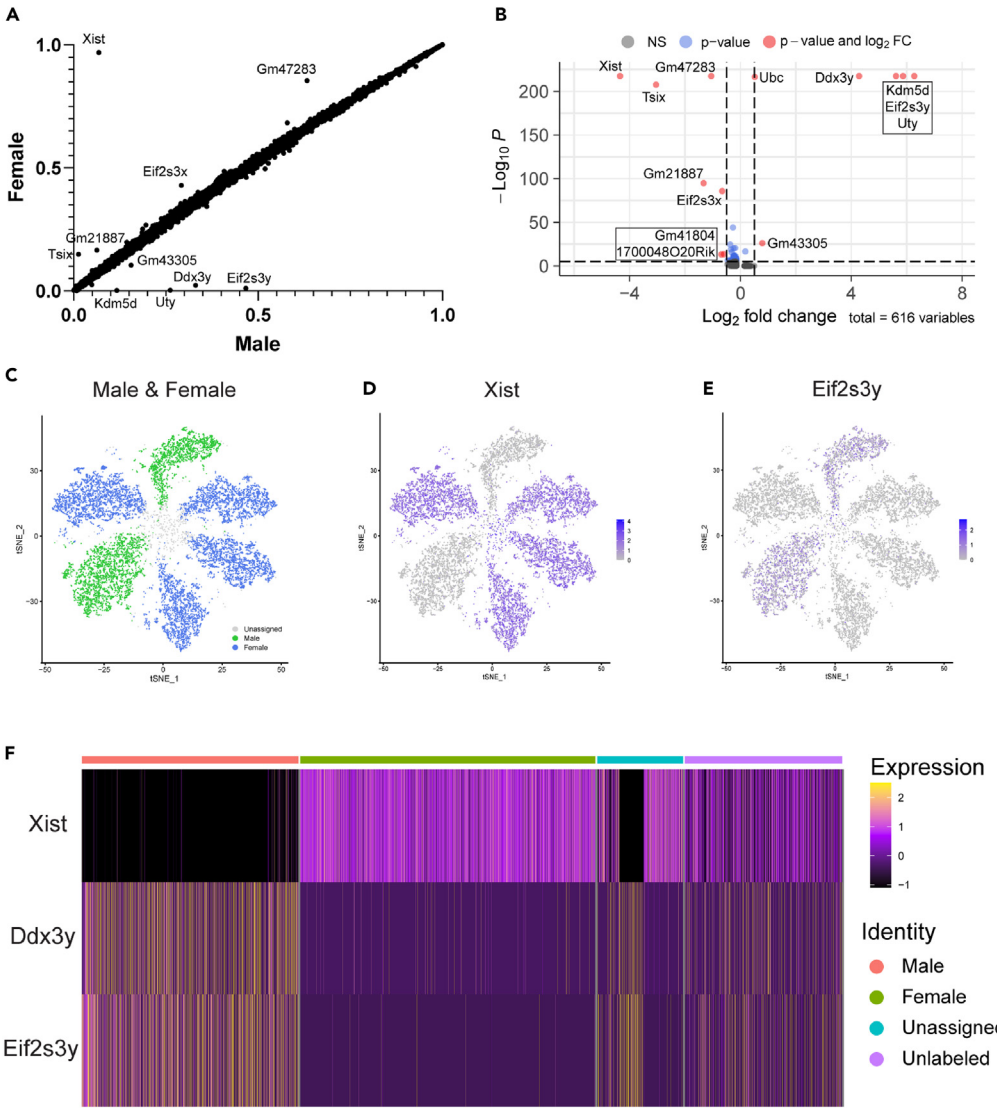


Figure 3. Sample barcoding reveals differential expression of sex-related genes

(A) All detected genes comparing cells from male and female samples plotted by percentage of cells with detected transcripts, excluding retinas from Exp3, which were male only.

(B) Volcano plot of top DEGs determined by the "bimod" test (FC, fold change; NS, not significant). Thirteen genes (red) had a >0.5 FC difference and a *p* value < 1e−05 between male and female cells, all of which were located on sex chromosomes.

(C) tSNE CMO feature plot labeled by sex of originating sample.

(D and E) X-linked (*Xist*) and Y-linked (*Eif2s3y*) genes display predicted expression patterns based on sample assignment.

(F) Heatmap of all experiments depicting scaled expression of three sex-linked genes segregated by assigned "Male" (red) or "Female" (green) identity. "Unlabeled Exp0" (purple), which pooled cells from four male and four female mice, was included for comparison (see also Table S3).

(*Erdr1x*) and *Gm47283* were also significant DEGs (Figures 3A and 3B). In contrast, the Unlabeled sample, which contained an equal number of male and female retinas, displayed an even distribution of sex-related genes (Figure 3F). It should be noted that the detection of sex-related gene expression varied between experiments (Table S3). While *Xist* was detected in most female-derived cells (95%–100%), *Ddx3y* and *Eif2s3y* were detected in fewer male-derived cells (20%–70%). A lower level of transcripts of sex-linked genes from the opposite sex was also detected in some cells (Table S3), suggestive of background transcript contamination. Overall, the specific enrichment of appropriate sex-linked genes is indicative that sample assignments were at least qualitatively accurate.

Accurate resolution of multiplets, transcriptional profiles derived from more than one cell, is a common challenge in scRNA-seq analysis. Leveraging our CMO-defined multiplets as ground truth, we removed multiplets in a two-stage clustering-based process. Following removal of low-quality cells (unique molecular identifier [UMI] ≤ 2,400, features ≤ 1,200, mitochondrial% > 10) and non-RGCs, we performed

unsupervised clustering based on transcriptomic similarity. We observed a large population of cells aggregated in the uniform manifold approximation and projection (UMAP) cluster plot (Cluster 0) (Figure S4A). This cluster was readily identifiable as a multiplet group because it primarily consisted of CMO-defined multiplets. The remaining cells in this cluster had multiplet-like qualities such as elevated UMI and non-specific expression of retinal cell class markers (mostly RGC/amacrine), and thus we designated them as multiplet-like cells (MLCs) (Figures S4B–S4D). These MLCs included cells from the unlabeled Exp0 but also cells that were "assigned" to a CMO. We expect "assigned" MLCs likely included multiplets where each cell was labeled with the same CMO, a combination of a labeled and unlabeled cell, or were otherwise misidentified. Multiplet clusters like Cluster 0 are frequently observed in scRNA-seq datasets and are routinely removed. Next, the remaining CMO-defined multiplets were distributed among RGC type clusters and would normally be challenging to identify (Figure S5A). We subclustered sets of transcriptionally similar RGC types to ask if these CMO-defined multiplets would aggregate with other MLCs. Indeed, both CMO-defined multiplets and MLCs co-clustered, enabling their identification (Figures S5B and S5C). We found that these multiplets/MLCs displayed non-specific expression of RGC subclass markers, suggesting they may consist primarily of RGC:RGC multiplets (Figures S5B‴ and S5C‴). Both MLCs and CMO-defined multiplets from subclustering analysis were removed as part of our QC pipeline (Figures 2A′–2C). With CMO-defined multiplets as a beacon, MLCs become more apparent. Notably, this approach enabled identification of MLCs from Exp0, which was not labeled with CMOs (Figures S5A–S5C). This indicates the feasibility of using a CMO-labeled reference dataset to curate multiplets from unlabeled samples of similar cellular composition.

Computational methods to detect multiplets in scRNA-seq data generally rely on features such as transcript count or mRNA heterogeneity, which can be biased if comparing cells with natural differences in these features. We asked how CMO-identified multiplets compare to multiplets identified by a commonly used computation method, DoubletFinder. We used a DoubletFinder prediction rate that would yield a similar number of multiplets. While there was a considerable overlap in labeling MLCs (Figure S5E), we found that spatially within UMAP clustering, DoubletFinder identified less regions populated by CMO-defined multiplets and had the tendency to label cells localized to clusters largely devoid of CMO-defined multiplets (Figures S5B, S5C and Table S4). Unlike multiplet calls by DoubletFinder, which have a hard UMI count threshold, CMO-defined multiplets and MLCs included some cells with lower UMI counts (Figure S5D). Overall, we conclude that CMO-based and DoubletFinder multiplet assignments display qualitative differences and our CMO approach may be more sensitive for the detection of multiplets from transcriptionally similar cells (such as RGC:RGC multiplets) and multiplets with lower UMI counts.

We next assessed RGC type distributions across CMO-labeled retinas. We overlaid sample assignments onto our clustered transcriptomic data (Figure 4A). We found that most clusters contained cells from every CMO-labeled retina at relatively even distributions, with the exceptions being relatively sparse RGC types (Table S5). "Unassigned" cells, for which sample identity was not assigned, demonstrated a similar distribution pattern (Figure 4A), suggesting these cells were qualitatively similar to "CMO-Assigned" cells. RGC frequencies also tracked closely with distributions in the RGC atlas[12] and Unlabeled cells (Figure 4B). RGC frequencies between experiments were highly consistent (Figures 4C, 4D and Table S5). These results demonstrated that CMO labeling does not bias collection toward certain RGC types and that RGC types exhibit predictable frequencies across biological replicates.

RGC gene expression across samples and cells

To determine whether cells from different retinas display gene expression differences, we analyzed global expression patterns on a per-retina basis (Figure 5). Comparison of gene expression from cells assigned to different retinas within each collection was highly similar (>0.97 Pearson correlation), even though each retina was dissected and processed separately (Figures 5A–5C). Cells from retinas collected in different experiments were also highly correlated (>0.77 Pearson correlation) but displayed more gene variance, which was likely due to batch effects (Figure 5C). We compared transcriptomic profiles between RGCs derived from left or right eyes from the same animal and detected only one DEG (*Gm19951*, a highly variable gene across samples) (Figure S6 and Table S6). Consistent with our analysis of sex-related gene expression in male and female samples, sex-linked genes were among the most variable genes across samples (Figures 5D, 5E and Table S6). There were additional genes such as *Lars2* or *Gm42418* with substantial levels of variation but no correlated phenotypic distinction. In general, global expression patterns among retinal samples were consistent within and between experiments. In addition, CMO labeling enabled the resolution of sex-specific differences, a valuable feature of the experimental design.

We next assessed whether type-specific gene expression patterns were consistently expressed across retinal samples. We analyzed the variance in gene expression of the type-defining markers (defined in Table1 and Figure S8A) between retinas and assessed whether variance was associated with cluster abundance (Figure S7 and Table S7). We found that variance in marker expression and cluster abundance were generally negatively correlated. Higher abundance types such as F_miniOFF and W3B had the lowest variability, while scarce types such as AlphaOFFT and AlphaONT were among the highest. Most types, including low abundance populations, recapitulated type-specific expression patterns across retinas with some specific exceptions (Table S7). This indicates that the chosen markers robustly identify RGC types, even on an individual retina basis. Overall, we demonstrate that CMO labeling enabled assessment of RGC type-specific gene expression in individual retinas but highlights a limitation for interrogation of expression patterns in low abundance types.

RGC atlas version 1.1

While not the primary objective of this study, our expanded RGC scRNA-seq dataset enabled a re-examination of RGC type annotations and expression patterns. The 47 clusters identified in our analysis correlated directly with the RGC atlas types with two exceptions. 16_ooDS_DV (on-off direction selective dorsal and ventral types) and 18_Novel each split into two clusters that we annotated as 16_ooDS_D, 16_ooDS_V, 18_Novel_a, and 18_Novel_b. The division of 16_ooDS_DV is consistent with Tran et al.,[12] who determined that this group contained two

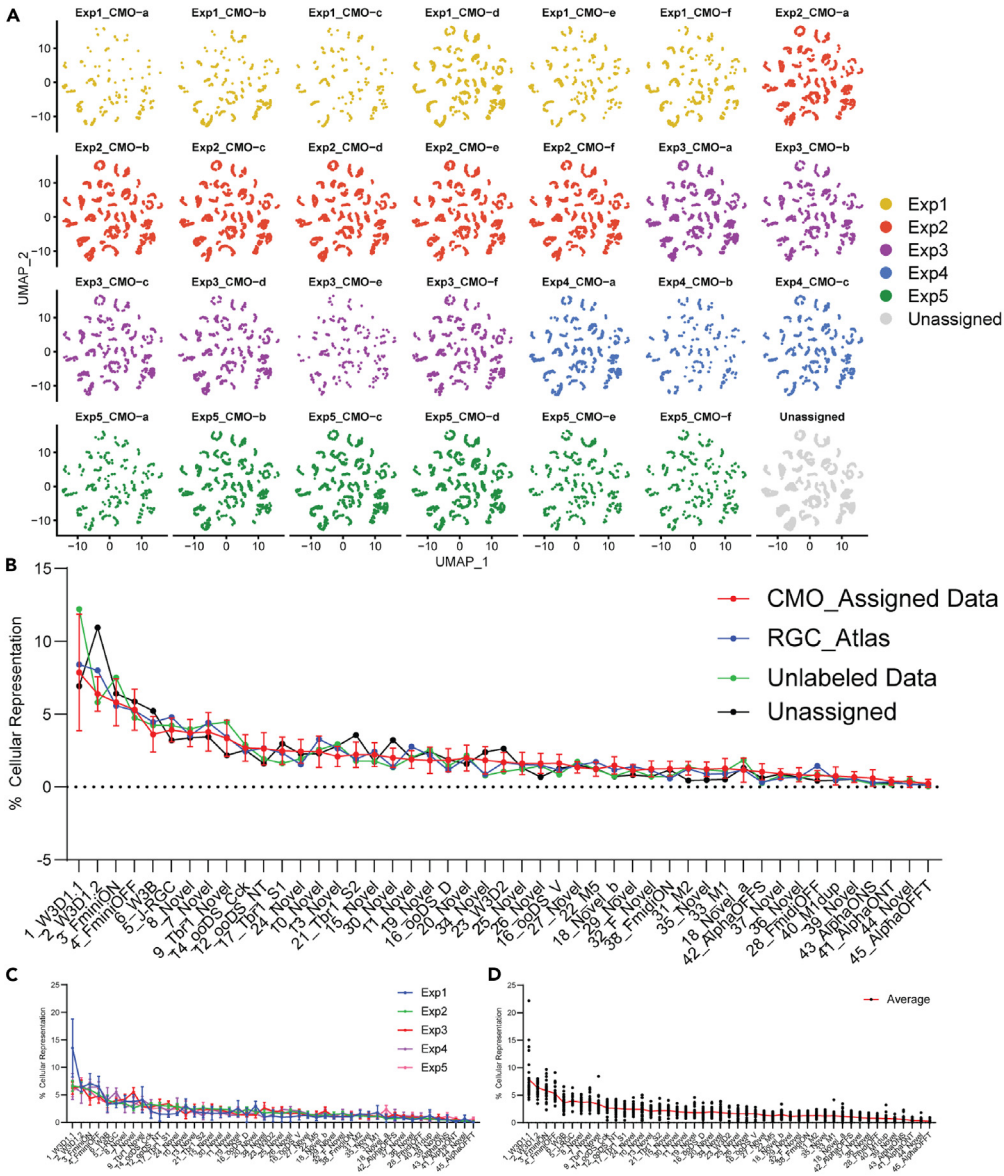


Figure 4. RGC type distribution is consistent across CMO-labeled retinas

(A) UMAPs showing distribution of RGCs from individual retinas across transcriptomic clusters (Exp0 not included). RGCs collected across CMO-labeled samples were evenly distributed across clusters. "Unassigned," cells not assigned to an individual retina by CMO labeling, were similarly distributed across clusters.

(B) Line plot depicting RGC type percent representation between the "RGC_Atlas" (Tran et al. 2019[12]), "Unlabeled Data" (Exp0), "CMO_Assigned Data" ($n = 27$), and the Unassigned data. Data are represented as mean +/− SD.

(C) Line plot of "CMO_Assigned Data" segregated by each experiment. Data are represented as mean +/− SD.

(D) Line plot of "CMO_Assigned Data" showing distribution of each individual retina per RGC type. Red line indicates average value. This graph indicates RGC type distribution was consistent across samples and labeling method and with previously published datasets. (see also Table S5).

transcriptionally similar types that could be manually subdivided based on the expression of *Calb1* and *Calb2*, established markers for the ooDS_D and ooDS_V types, respectively.[20] Indeed, the ooDS_D and ooDS_V clusters were specifically enriched for *Calb1* or *Calb2* (Figures 6B and 6B''). 18_Novel_a and 18_Novel_b were distinguished by the expression of *Pcdh11x* and *Pcdh20/Prkg2*, among other markers (Figures 6C, 6C'', S8 and Table 1). 18_Novel, along with other "Novel" types described in Tran et al. 2019,[12] was molecularly defined but has not been completely characterized. However, annotation of these four types was independently verified in a large-scale integrated analysis of mouse retinal scRNA-seq datasets.[21] In addition, we analyzed our clustered dataset to validate and revise cell-type-specific markers. We identified markers that designate transcriptomically related groups of RGCs including groups defined by the expression of specific gene markers like *Irx3* and *Bnc2* or co-expression of *Mafb* and *Kcnd2* (Table 1 and Figure S8). We defined a minimal set of molecular markers that identify

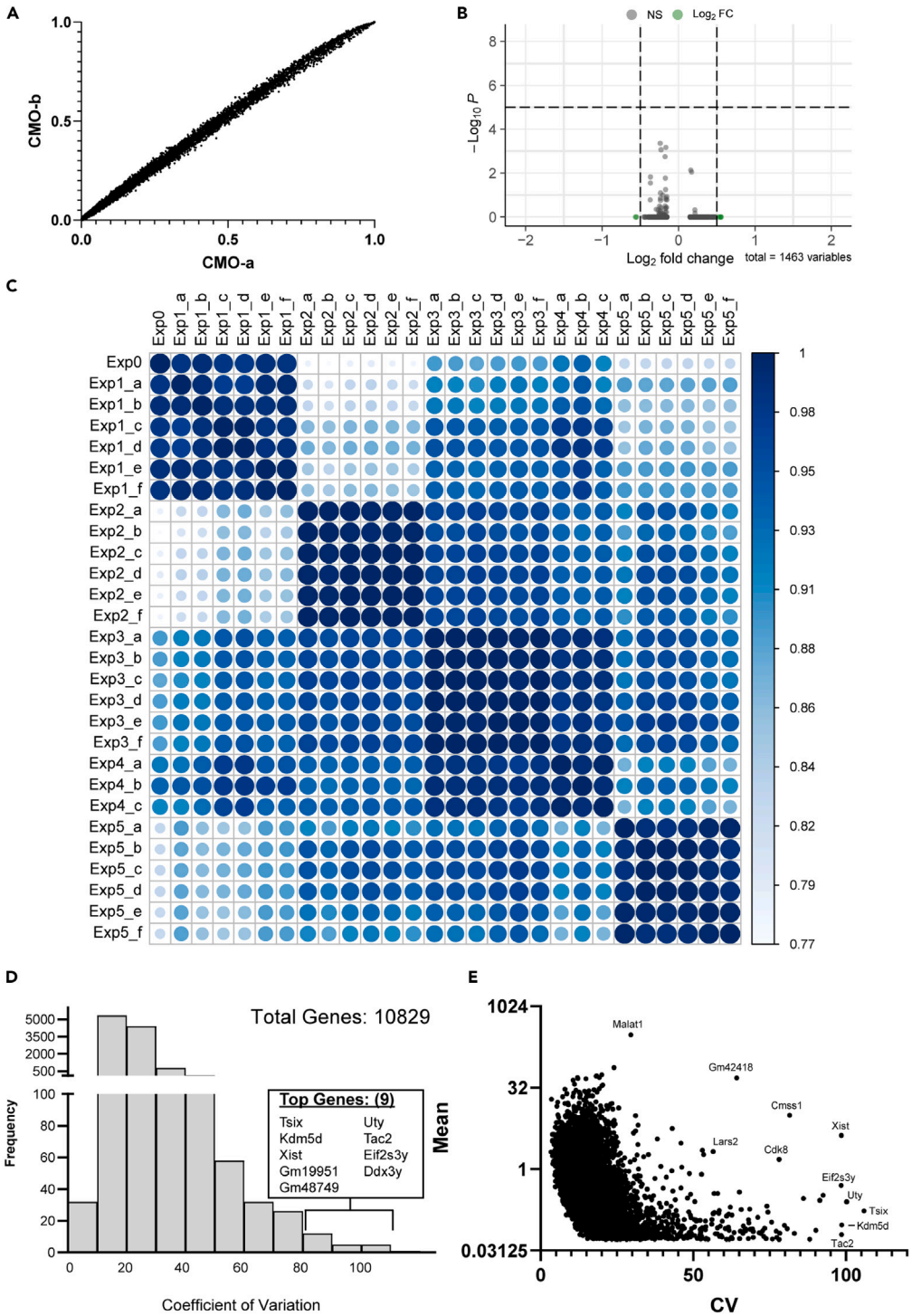


Figure 5. Gene expression patterns are highly correlated between CMO-labeled retinas

(A and B) A representative comparison showing the low transcriptional variance between cells assigned to different CMOs within each collection. No significantly different genes were identified by the "bimod" test comparing cells assigned to CMO-a and CMO-b from Exp2.

(C) Correlation analysis (Pearson) comparing the average gene expression of the top 5,000 variable genes between all retinas. Correlations ranged from 0.77–1.00 across all samples and 0.97–1.00 comparing samples from the same experiment, indicating high transcriptional similarity among cells derived from different retinas within and between experiments.

Figure 5. *Continued*

(D) Histogram showing distribution of the coefficient of variation (CV, standard deviation divided by the mean) across all detected genes (scaled mean > 0.05). The 9 most variable genes (>80 CV) are listed.

(E) Scatterplot comparing CV to mean expression level for all detected genes (see also Figure S7 and Table S6).

each type, requiring no more than five markers for each (Table 1). The combined expression of our updated markers was highly type specific and sufficient to resolve all adult RGC types (Figure S8). Lastly, we curated annotations from two recent studies that used patch-clamp RNA sequencing (Patch-seq) to associate molecular types with functional characteristics (Table 1).[22,23] Our updated RGC annotations improved type resolution and provide useful information for identifying or gaining genetic access to specific types.

DISCUSSION

We have developed an RGC scRNA-seq protocol that incorporates sample multiplexing using CMO-based barcoding. Multiplexing did not affect data quality, cell type prediction, or expression patterns. This dataset enabled evaluation of RGC cell type distributions and expression patterns from individual retinas. Sample multiplexing has several benefits for studying gene expression in rare cell populations including increasing the number of biological replicates and directly matching transcriptomic profiles with sample phenotypes and other metadata like sex or age. Implementation of these approaches enhances the efficiency and scalability of testing different experimental conditions. It is important to note that there are several methods for sample multiplexing for scRNA-seq including antibody-based methods such as Cite-seq,[6] or methods which target the genome such as CellTag Indexing.[9] We selected the CMO-labeling approach for this study because this method is readily integrated with our collection protocols. However, the optimal multiplexing approach will differ by experiment.

Using CMO-based multiplexing, we demonstrated a sample assignment rate of ~82%, ranging from 67% to 94% across samples, with an average recovery of 1,373 RGCs/retina (excluding Exp0 and Exp1) post-QC. Subjectively, we found that assignment rate may correlate with sample quality, as samples in which more cells were filtered out by standard QC metrics also had higher fractions of unassigned cells. Within genotypically homogenous samples like tissues from congenic mice, it is challenging to independently verify classification accuracy. We utilized sex-related genes to estimate cross-sample "contamination" on a qualitative level. In general, appropriate enrichment of sex-specific genes was observed across samples, but notably the detection level (e.g., percent of cells expressing each sex-linked marker) and background contamination were variable across collections. While informative, evaluation of sex-related genes only divided samples into two groups for each collection, whereas six retinas were assessed. Therefore, additional sample-specific features would be required to quantitatively evaluate cross-sample contamination. Overall, we conclude that CMO labeling efficiently and accurately identified retinal origin for most cells.

An additional useful feature of CMO labeling was the ability to accurately identify multiplets and, in conjunction, MLCs. Multiplets are present in most scRNA-seq datasets and can impact clustering and confound interpretation if not properly resolved.[24–26] Bioinformatic methods are commonly used to identify multiplets within scRNA-seq data[27–29]; however, these rely on gross transcriptomic metrics such as transcripts or UMIs per cell, which can differ by cell type. Without a secondary method, it is difficult to verify the accuracy of computationally derived multiplet detection. Our analysis demonstrated that multiplet detection differed substantially between CMO-based assignments and the widely used computational method DoubletFinder.[29] By identifying MLCs that co-cluster with CMO-defined multiplets, our approach resolved MLCs with lower UMI/transcript counts and may be more accurate for detecting multiplets of transcriptionally similar cells (e.g., RGC:RGC doublets). Notably, our CMO-labeled data readily enabled detection of MLCs in an unlabeled sample, suggesting our dataset could be used as a reference for multiplet detection in pooled RGC scRNA-seq collections. A limitation to CMO multiplet assignments is that precision inherently scales with the number of samples labeled. For example, our experiments assessed 6 CMO-labeled retinas per experiment, so theoretically ~83% of doublets would be expected to have two or more different CMO tags, while ~17% would have the same tag, assuming equal cell input and labeling efficiency. We conclude that sample multiplexing is advantageous for the detection of cell multiplets and its utility extends to unlabeled samples. However, it may still be advisable to consider additional computation methods for independent verification.

The primary purpose of incorporating sample multiplexing into scRNA-seq pipelines is to resolve type-specific transcriptional differences in individual samples. Global and type-specific expression patterns were highly consistent across samples. We found little expression differences when comparing cells from "left" and "'right" eyes. In contrast, sex-related genes demonstrated appropriate enrichment patterns based on predicted sample origin. Type-specific markers were generally consistent across samples, but these patterns were less well resolved for "low abundance" types. The top 5 scarcest clusters had an average of 5.2 cells/retina, 5 median clusters had an average of 23.1 cells/retina, and the top 5 most abundant clusters had an average of 66.1 cells/retina. Thus, appropriate caution should be used in the assessment of "low abundance" types in individual samples, but overall type-specific expression differences across retinas were readily detectable. Our results also demonstrated that RGC types were evenly distributed across samples, supporting the conclusion that these molecular groupings represent cell types, rather than cell states or batch-related effects. We found that RGC types and their expression patterns are generated with remarkable consistency across individual retinas, highlighting the precise developmental regulation of type specification in this complex neural tissue. As scRNA-seq is increasingly used to evaluate transcriptomic shifts in disease and development, sample multiplexing is a useful tool to improve the precision of scRNA-seq analysis and enhance testing across experimental conditions and biological replicates, particularly for rare cell populations.

Limitations of the study

Our study concludes that CMO-based multiplexing enhances precision of scRNA-seq with few drawbacks. However, there are limits to its applicability. The labeling efficiency can be variable, and it may be difficult to empirically determine the precise accuracy and efficiency of

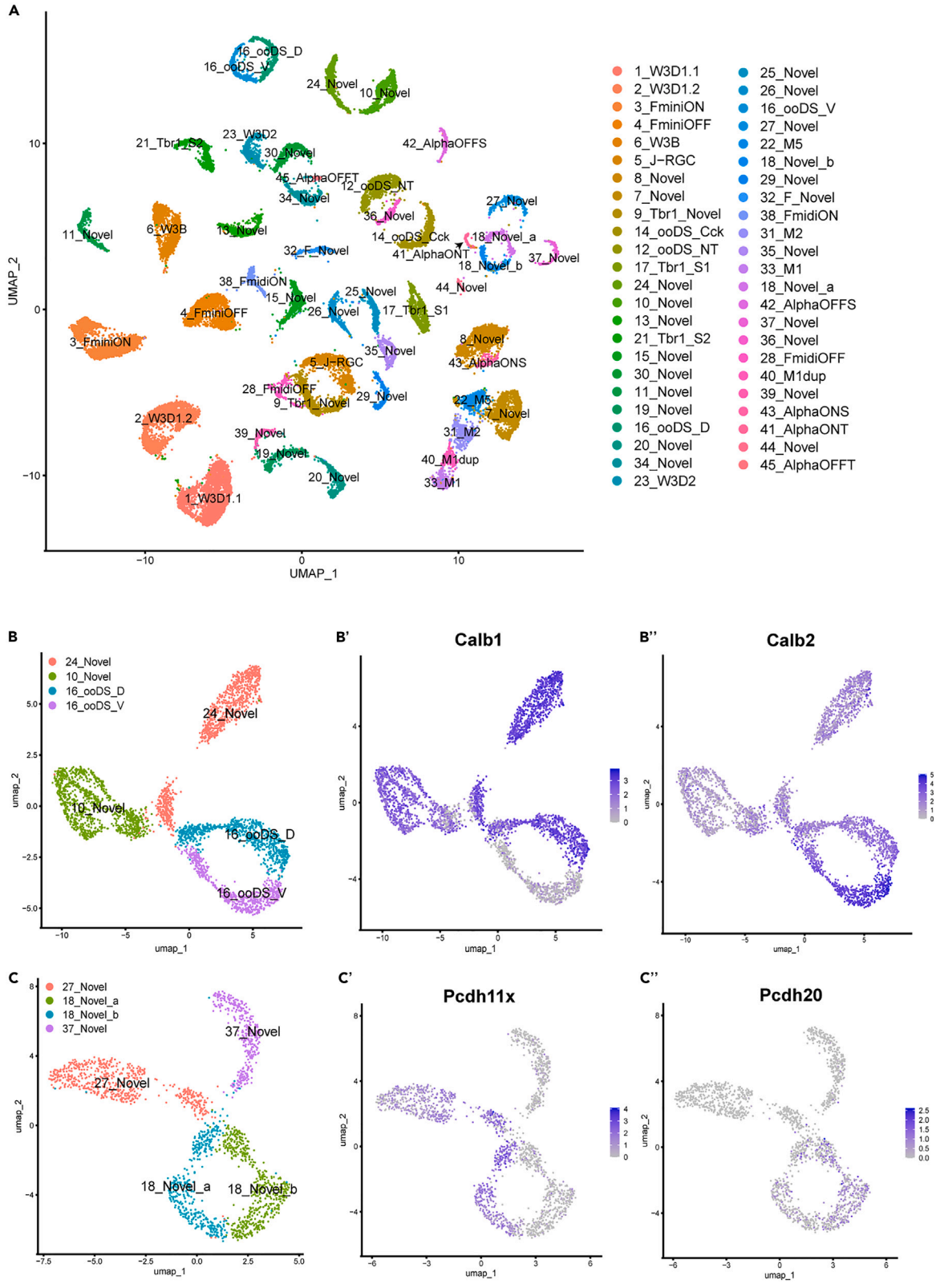

A
UMAP_1
UMAP_2
1_W3D1.1
2_W3D1.2
3_FminiON
4_FminiOFF
6_W3B
5_J−RGC
8_Novel
7_Novel
9_Tbr1_Novel
14_ooDS_Cck
12_ooDS_NT
17_Tbr1_S1
24_Novel
10_Novel
13_Novel
21_Tbr1_S2
15_Novel
30_Novel
11_Novel
19_Novel
16_ooDS_D
20_Novel
34_Novel
23_W3D2
25_Novel
26_Novel
16_ooDS_V
27_Novel
22_M5
18_Novel_b
29_Novel
32_F_Novel
38_FmidiON
31_M2
35_Novel
33_M1
18_Novel_a
42_AlphaOFFS
37_Novel
36_Novel
28_FmidiOFF
40_M1dup
39_Novel
43_AlphaONS
41_AlphaONT
44_Novel
45_AlphaOFFT
B
B'
Calb1
B''
Calb2
C
C'
Pcdh11x
C''
Pcdh20
umap_1
umap_2

Figure 6. Refinement of the RGC atlas annotation

(A) UMAP of the transcriptomic data with RGC type annotations.

(B) Feature plots depicting enrichment of *Calb1* and *Calb2* expression in ooDS_V and ooDS_D, respectively.

(C) Feature plots depicting enrichment of *Pcdh11x* and *Pcdh20* expression in 18_Novel_a and 18_Novel_b, respectively. (See also Figure S8).

labeling in an experiment. Identification of multiplets is improved by analysis of CMO barcodes but remains imperfect, though this was partially circumvented by focusing on multiplet clusters rather than individual cells. Our study included samples from mice ranging from 7 to 12 weeks old. However, our experimental design was not optimized to study age-dependent effects because retinas from each age were collected in separate batches and represent a limited adult age range. Sample multiplexing would be useful in future studies on age-dependent effects, since samples from multiple ages could be compared within a single collection. In addition, despite enabling resolution of RGCs across individual samples, the capacity to assess relatively low-abundance RGC types on a per-retina basis is limited, as demonstrated by the increased variance in marker gene expression.

RESOURCE AVAILABILITY

Lead contact

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request. Further information and requests for resources or reagents should be directed to and will be fulfilled by Nicholas M Tran (Nicholas.Tran@bcm.edu).

Materials availability

This study did not generate any new reagents.

Data and code availability

- Data: single-cell RNA-seq data (raw and processed) have been deposited to the gene expression omnibus along with the appropriate metadata and are publicly available. Accession number is listed in the key resources table.
- Code: all original relevant code has been cited. Our workflow code is stored as a GitHub repository and is publicly available. Hyperlink is available in the key resources table.
- Additional Information: any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was performed using the support of these funding sources: NIH R00EY029360, Whitehall Foundation, TIRR Foundation Mission Connect Grant to N.M.T., NIH R01EY025601, NIH R21EY033458, Research to Prevent Blindness (Unrestricted Grant), and the Retina Research Foundation. In addition, scRNA-seq library preparation was performed by the Baylor College of Medicine Advanced Technology Core Labs Single Cell Genomics Core, which is partially supported by NIH S10OD025240 and CPRIT RP200504.

AUTHOR CONTRIBUTIONS

All authors have read and approved the final version of this manuscript and meet the Journal's criteria for authorship. All experiments were performed in the laboratories of N.M.T., B.J.F., G.M., and R.C. by J.M., N.M.T., T.K.C., M.P.P., Y.P., and Y.L. The study was conceived and designed by J.M., Y.P., N.M.T., B.J.F., and G.M. Data were analyzed and interpreted by J.M., N.M.T., B.J.F., and G.M. The manuscript was drafted and revised by J.M., N.M.T., B.J.F., and G.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODELS
- METHOD DETAILS
  - Cell preparation and sequencing
  - Bioinformatics
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111250.

Received: April 13, 2024
Revised: August 9, 2024
Accepted: October 22, 2024
Published: October 24, 2024

REFERENCES

1. Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. *9*, 1–12. https://doi.org/10.1186/s13073-017-0467-4.
2. Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med. *50*, 1–14. https://doi.org/10.1038/s12276-018-0071-8.
3. Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., and Luo, Y. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. Clin. Transl. Med. *12*, e694. https://doi.org/10.1002/ctm2.694.
4. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214. https://doi.org/10.1016/j.cell.2015.05.002.
5. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201. https://doi.org/10.1016/j.cell.2015.04.044.
6. Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., 3rd, Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224. https://doi.org/10.1186/s13059-018-1603-1.
7. McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat. Methods *16*, 619–626. https://doi.org/10.1038/s41592-019-0433-8.
8. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94. https://doi.org/10.1038/nbt.4042.
9. Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G.C., Yang, X., Kirita, Y., and Morris, S.A. (2019). CellTag Indexing: genetic barcode-based sample multiplexing for single-cell genomics. Genome Biol. *20*, 90. https://doi.org/10.1186/s13059-019-1699-y.
10. Cheng, J., Liao, J., Shao, X., Lu, X., and Fan, X. (2021). Multiplexing Methods for Simultaneous Large-Scale Transcriptomic Profiling of Samples at Single-Cell Resolution. Adv. Sci. *8*, 2101229. https://doi.org/10.1002/advs.202101229.
11. Jacobi, A., Tran, N.M., Yan, W., Benhar, I., Tian, F., Schaffer, R., He, Z., and Sanes, J.R. (2022). Overlapping transcriptional programs promote survival and axonal regeneration of injured retinal ganglion cells. Neuron *110*, 2625–2645.e7. https://doi.org/10.1016/j.neuron.2022.06.002.
12. Tran, N.M., Shekhar, K., Whitney, I.E., Jacobi, A., Benhar, I., Hong, G., Yan, W., Adiconis, X., Arnold, M.E., Lee, J.M., et al. (2019). Single-Cell Profiles of Retinal Ganglion Cells Differing in Resilience to Injury Reveal Neuroprotective Genes. Neuron *104*, 1039–1055.e12. https://doi.org/10.1016/j.neuron.2019.11.006.
13. Jacobi, A., and Tran, N.M. (2023). Defining Selective Neuronal Resilience and Identifying Targets for Neuroprotection and Axon Regeneration Using Single-Cell RNA Sequencing: Experimental Approaches. Methods Mol. Biol. *2636*, 1–18. https://doi.org/10.1007/978-1-0716-3012-9_1.
14. Jeon, C.J., Strettoi, E., and Masland, R.H. (1998). The major cell populations of the mouse retina. J. Neurosci. *18*, 8936–8946. https://doi.org/10.1523/jneurosci.18-21-08936.1998.
15. Dräger, U.C., and Olsen, J.F. (1981). Ganglion cell distribution in the retina of the mouse. Invest. Ophthalmol. Vis. Sci. *20*, 285–293.
16. Claes, M., and Moons, L. (2022). Retinal Ganglion Cells: Global Number, Density and Vulnerability to Glaucomatous Injury in Common Laboratory Mice. Cells *11*, 2689. https://doi.org/10.3390/cells11172689.
17. Park, Y.H., Snook, J.D., Zhuang, I., Shen, G., and Frankfort, B.J. (2020). Optimized culture of retinal ganglion cells and amacrine cells from adult mice. PLoS One *15*, e0242426. https://doi.org/10.1371/journal.pone.0242426.
18. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
19. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *16*, 14049. https://doi.org/10.1038/ncomms14049.
20. Kay, J.N., De la Huerta, I., Kim, I.J., Zhang, Y., Yamagata, M., Chu, M.W., Meister, M., and Sanes, J.R. (2011). Retinal ganglion cells with distinct directional preferences differ in molecular identity, structure, and central projections. J. Neurosci. *31*, 7753–7762. https://doi.org/10.1523/jneurosci.0907-11.2011.
21. Li, J., Choi, J., Cheng, X., Ma, J., Pema, S., Sanes, J.R., Mardon, G., Frankfort, B.J., Tran, N.M., Li, Y., and Chen, R. (2024). Comprehensive single-cell atlas of the mouse retina. iScience *27*, 109916. https://doi.org/10.1016/j.isci.2024.109916.
22. Huang, W., Xu, Q., Su, J., Tang, L., Hao, Z.Z., Xu, C., Liu, R., Shen, Y., Sang, X., Xu, N., et al. (2022). Linking transcriptomes with morphological and functional phenotypes in mammalian retinal ganglion cells. Cell Rep. *40*, 111322. https://doi.org/10.1016/j.celrep.2022.111322.
23. Goetz, J., Jessen, Z.F., Jacobi, A., Mani, A., Cooler, S., Greer, D., Kadri, S., Segal, J., Shekhar, K., Sanes, J.R., and Schwartz, G.W. (2022). Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression. Cell Rep. *40*, 111040. https://doi.org/10.1016/j.celrep.2022.111040.
24. Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J.C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. Nat. Methods *14*, 565–571. https://doi.org/10.1038/nmeth.4292.
25. Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145. https://doi.org/10.1038/nrg3833.
26. Kiselev, V.Y., Andrews, T.S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. Nat. Rev. Genet. *20*, 273–282. https://doi.org/10.1038/s41576-018-0088-9.
27. DePasquale, E.A.K., Schnell, D.J., Van Camp, P.J., Valiente-Alandí, Í., Blaxall, B.C., Grimes, H.L., Singh, H., and Salomonis, N. (2019). DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. Cell Rep. *29*, 1718–1727.e8. https://doi.org/10.1016/j.celrep.2019.09.082.
28. Bais, A.S., and Kostka, D. (2020). scds: computational annotation of doublets in single-cell RNA sequencing data. Bioinformatics *36*, 1150–1158. https://doi.org/10.1093/bioinformatics/btz698.
29. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. *8*, 329–337.e4. https://doi.org/10.1016/j.cels.2019.03.003.
30. Vong, L., Ye, C., Yang, Z., Choi, B., Chua, S., Jr., and Lowell, B.B. (2011). Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. Neuron *71*, 142–154. https://doi.org/10.1016/j.neuron.2011.05.028.
31. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140. https://doi.org/10.1038/nn.2467.
32. Chen, T., and Guestrin, C. (2016). In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery), pp. 785–794.
33. Dempster, A.P., Nan, M.L., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. J. Roy. Stat. Soc. *39*, 1–38.
34. Shekhar, K., Whitney, I.E., Butrus, S., Peng, Y.R., and Sanes, J.R. (2022). Diversification of multipotential postmitotic mouse retinal ganglion cell precursors into discrete types. Elife *11*, e73809. https://doi.org/10.7554/eLife.73809.
35. McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467. https://doi.org/10.1093/bioinformatics/bts714.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat anti-mouse CD73	Becton Dickinson	Cat: #550738; RRID: AB_393857
CD90.2 (Thy-1.2) Monoclonal Antibody, APC	Invitrogen	Cat: 17-0902-83; RRID: AB_469422
Reagents		
Tris-HCl	Sigma-Aldrich	T6066
Papain	Worthington	#LS003126
Ovomucoid	Worthington	3086-39b11094
Actinomycin D	Sigma-Aldrich	A1410
Bovine Serum Albumin	Sigma-Aldrich	A-4161
Ames' Media	Sigma-Aldrich	A1420
L-Cysteine	Sigma-Aldrich	C7477
DNase I	Worthington	#LS002007
10x Next GEM Single Cell 3′ Reagent Kit v3.1 with Feature Barcoding Technology for Cell Multiplexing	10x Genomics	PN-1000268, PN-1000261, PN-1000243
Mouse Lines		
Vglut2-Cre	Vong, L. et al. 2011[30]	Jax #028863
Ai9-Tdtomato	Madisen, L. et al. 2010[31]	Jax #007909
Deposited data		
Raw and processed data	This paper	GEO: GSE254587
Software and algorithms		
Cell Ranger 6.1.2	Zheng, G. X. Y. et al. 2017[19]	https://www.10xgenomics.com/support/software/cell-ranger/latest
Seurat 4.0	Hao, Y. et al. 2021[18]	https://satijalab.org/seurat/
HTODemux	Stoeckius, M et al. 2018[6]	https://satijalab.org/seurat/articles/hashing_vignette.html
Xgboost	Chen and Guestrin, 2016[32]	https://cran.r-project.org/web/packages/xgboost/index.html
DoubletFinder	McGinnis, C. S. et al. 2019[29]	https://github.com/chris-mcginnis-ucsf/DoubletFinder
Other		
Code Repository for this Publication	This paper	https://github.com/NMTran/NMTranLab/tree/main/Ma_CMO-Labeling_of_RGCs

EXPERIMENTAL MODELS

We used *Vglut2-Cre* (Jax# 028863[30]) mice mated to the Cre-dependent *Ai9-Tdtomato* (Jax# 007909[31]) reporter line, to fluorescently label RGCs. All animals were treated in accordance with the National Institutes of Health (NIH) guidelines, the Association for Research in Vision and Ophthalmology (ARVO) Statement for Use of Animals in Ophthalmic and Vision Research, and the Baylor College of Medicine Institutional Animal Care and Use Committee (IACUC) welfare guidelines. All experiments were carried out using mice from 7 to 12 weeks of age with inclusion of both sexes (Table S1). In addition, sex was an experimental variable of our study.

METHOD DETAILS

Cell preparation and sequencing

Retinal dissociation, RGC enrichment, and CMO labeling

Retinas were dissociated as previously described with some modifications.[12,13,17] Mice were euthanized and retinas were dissected out in oxygenated Ames' media (A1420, Sigma-Aldrich, St. Louis, MO) in the presence of Actinomycin D (ActD; A1410, Sigma-Aldrich; 30 μM) along with the removal of ciliary bodies, vitreous, and inner limiting membranes (ILM). Up to six retinas were used per experiment and processed

independently. Retinas were digested in papain solution (1.4% Papain (#LS003126, Worthington, Lakewood, NJ), 0.1% DNaseI (#LS002007, Worthington), 0.1% L-Cysteine (C7447, Sigma-Aldrich)) for 15 min and dissociated into single cells by gentle manual trituration using a pipette in ovomucoid solution (0.3% Ovomucoid (#3086-39b10094, Worthington), 0.3% BSA (A4161, Sigma-Aldrich), 0.05% DNaseI in oxygenated Ames'). Dissociated retinal cells were negative-immunopanned using anti-mouse CD73 (#550738; BD Pharmingen, San Jose, CA; 0.20% diluted in 50mM Tris-HCl pH 9.5 (T6066, Sigma-Aldrich)) to deplete rod photoreceptors. The remaining cells are resuspended in our working cell buffer solution (Ames'/0.4% BSA +3 μM ActD) and incubated with CD90.2/Thy1.2 antibody conjugated to APC (#17-0902-83; Invitrogen, Carlsbad, CA; 1 μL per 5 million cells) for 15 min to label RGCs. Cells were then washed and labeled with CMO's using the 10x Next GEM Single Cell 3' Reagent Kit v3.1 with Feature Barcoding Technology for Cell Multiplexing (PN-1000268, PN-1000261, PN-1000243). Each retina-derived cell suspension is mixed with a unique CMO at baseline concentration and incubated for exactly 5 min and subsequently the reaction is paused on ice. Each cell suspension was then washed by resuspension in Ames'/1% BSA +3 μM ActD followed by centrifugation. CMO-labeled samples were pooled, counterstained with DAPI to identify non-viable cells, and sorted using FACS. RGCs were purified by FACS based on co-expression of CD90.2/Thy1.2 (APC) and Vglut2 (TdTomato).

Single-cell sequencing

Cells are resuspended in Ames'/0.4% BSA, triturated into a single-cell suspension, and run on the Chromium Controller (10x Genomics). Library construction was performed as per the manufacturer's protocol (10x Genomics). Library construction is performed separately for gene expression and CMOs. Approximately 10,000–25,000 cells were loaded per channel. Libraries were sequenced using the Illumina NovaSeq 6000 with paired-end sequencing (GEX: TruSeq Read1, TruSeq Read2; CMO: Nextera Read1, Nextera Read2).

Bioinformatics

Cell Ranger

Sequencing Alignment was done using the Cell Ranger 6.1.2 mkfastq and multi pipelines and introns were retained. Cell Ranger based demultiplexing was performed but was not the primary demultiplexing method used for assessment of cell to retina assignment. Sequences were aligned to the GRGCm38/mm10-2020A genome reference build to generate cell-by-gene gene expression matrix. A cell-by-CMO matrix for CMO sample multiplexing was independently generated.

Computational methods to define RGC collection

Gene expression matrices were processed using Seurat 4.0 in R-programming language. Our cumulative RGC dataset includes six collections (Exp0 was Unlabeled, Exp1-5 were CMO-labeled). The data was curated by a series of QC filters resulting in a total RGC number of 41,782 cells. For QC, we initially removed cells with poor quality transcriptomes (mitochondrial genes >10%, UMIs $\leq$ 2400, features $\leq$ 1200) followed by non-RGC clusters based on marker expression. Subsequently, we performed targeted removal of multiplets. This initially began with removing a large multiplet cluster predominantly of RGC/non-RGC multiplet origin. Next, we identified MLC clusters from RGC subgroups (subclusters of similar RGC-types) by labeling clusters of cells highly populated with CMO-defined multiplets. Both CMO-defined multiplets and MLCs were removed as part of the QC pipeline. Data was processed as follows: the data was first normalized and scaled using the functions of the Seurat package.[18] Then, randomized PCA analysis was used for dimensional reduction accounting for 100 principal components. Then, we performed pattern classification by calculating the k-nearest neighbors graph (Euclidean distance) followed by clustering using the Louvain algorithm (k. param = 30).

Computational methods for demultiplexing

From Cell Ranger, the CMO multiplex pipeline generated a CMO counts matrix as well as assignment-confidence levels calculating using Cell Ranger's tag assignment algorithm. The CMO counts matrix was similar to the traditional gene expression matrix, but the features were the CMO counts per cell. We utilized two methods to assign CMO labels, Cell Ranger (10x Genomics) and HTOdemux.[6] Each method was used to identify cells as assigned (labeled with one CMO), unassigned (no CMO label), or a multiplet (two or more CMO labels). Cell Ranger's tag assignment algorithm was based on a customized Expectation Maximization algorithm[19,33] to fit a binomial distribution. The algorithm assumes a Gaussian distribution of counts containing background noise and actual counts from cell tags. Cells with a confidence threshold >0.9 were assigned. The HTODemux algorithm utilizes k-medoid clustering on the normalized CMO counts and fits a negative binomial distribution for each CMO. Barcodes with CMO signals >0.99 quantile was considered "positive" for that CMO. As depicted in Figure 2B, we found that this method was more stringent in classifying multiplets in our dataset. For cluster visualization, the CMO counts matrix was processed as a Seurat object following a standard workflow and visualized by tSNE distribution.

RGC validation

To validate our clustering of RGCs, we compared our transcriptomic data to that of the published RGC Atlas using a multi-class classifier (R package Xgboost)[32] following the methods outlined in Shekhar et al., 2022.[34] The model was trained on the RGC Atlas clusters were tested against clusters of our dataset.

Multiplet comparison

For comparing DoubletFinder and CMO-based multiplets, we analyzed our dataset after the initial removal of the "RGC/non-RGC multiplet cluster". DoubletFinder was performed using the associated R package[29] (sctransform not used), accounting for 50 principal components, and assumed a 15% doublet rate.

DEG analysis

Top DEGs are determined by using the FindMarkers function as part of the R Seurat package using the "bimod" test.[35] We evaluated genes that were expressed in >5% of cells and set an initial fold change threshold of >0.15. Additional filters for fold change and *p*-value were applied for specific figures and tables. Retina-wide comparisons were evaluated using averaged expression levels (for each gene, expression was averaged across all cells per retina). Similarity of pseudo-bulk retinal transcriptomes was assessed using Pearson correlation between the average expression of the top 5000 variable genes between all 27 CMO-labeled retinas. Gene variance was assessed by calculating the coefficient of variation (CV, standard deviation divided by the mean) of each gene between all 27 retinas. For top gene variance rankings, we removed lowly expressed genes (scaled average < 0.05) of which CV calculations may be inaccurate. Cluster-specific gene variance is similarly assessed using CV but applied for each cluster on only the associated marker genes. For cluster-specific variance calculations, retinas from Exp1 were not included because of low cell representation and retinas without cells belonging to a particular cluster were not factored in for the data of that cluster.

QUANTIFICATION AND STATISTICAL ANALYSIS

All descriptions of quantification, sample sizes, and/or parameters used are specified in the corresponding figure legends. In all cases, n represents biological replicates. All error bars are representative of standard deviation. For boxplots, the whiskers represent the min and max and box lines represent the 25^{th} percentile, median, and 75^{th} percentile, respectively. Pearson correlation was used to determine transcriptomic similarity between samples. Software used is in key resources table. Coefficient of variation was calculated by taking the standard deviated divided by the mean. The "bimod" statistical test was used for differential gene analysis. No other statistical tests were performed in this study.