

Single Cell Expression Analysis for Understanding the Development of Glaucoma

By Alex Leung

*Thesis submitted to University College London for the degree of
Doctor of Philosophy*

2022

UCL Institute of Ophthalmology
Faculty of Brain Sciences

Declaration

I, Alex Leung, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. This thesis does not exceed 100,000 words.

Abstract

Glaucoma is characterized as a group of eye diseases where the progressive damage of neurons, particularly Retinal Ganglion Cells (RGCs), leads to vision loss. This disease affects more than 70 million people worldwide, with approximately 10% being bilaterally blind, making it the leading cause of irreversible blindness in the world. The initiation and progression of the disease is still unknown, but studies have suggested the involvement of particular cell types in the retina that relate to the pathogenesis of glaucoma. Single cell RNA sequencing (RNA-seq) analysis is a new technology that provides insight into the gene expression profiles of different cell types. In this study, we employed it to elucidate the transcriptomic changes in various cell types during glaucoma progression.

ABCA1^{-/-} mice were used as a normal tension glaucoma model. Single cell RNA-seq experiments were conducted on three wild type (WT) and five knockout (KO) retinal tissues. The data of 62,479 cells were integrated and major cell types were identified, including Müller glia, astrocytes, microglia and RGCs. Ontological analysis suggested strong activation of neuroinflammation and senescence related pathways in KO samples, with specific pathways identified affecting certain cell types. Evidence of macrophage invasion further suggests a knockout-induced inflammatory response, accompanied by sub-type specific RGC degeneration due to excitotoxicity.

P2Y6^{-/-} mice were used as a high intraocular pressure (IOP) glaucoma model. 105,772 cells from three WT and three KO retinal tissues were analysed using single cell RNA-seq, with major cell types identified such as RGCs and glial cells. Neuroinflammation and senescence pathways activation was again observed, along with angiogenesis, hypoxia and fibrosis activities activated in knockout glial population.

Overall, our study utilized the power of single cell analysis to improve our understanding of the roles and involvement of different cell types under different glaucoma development and progression pathways. We propose these specific cell types and mechanisms could be involved in the development of glaucoma

pathogenesis, thus provided data to support future interests in developing potential therapeutic targets in the area.

Impact Statement

Glaucoma is a disease which affects more than 70 million people worldwide, with approximately 10% being bilaterally blind, making it the leading cause of irreversible blindness in the world. Although some risk factors have been identified to increase the likelihood of developing glaucoma (such as age, increase in intraocular pressure, family history and genetics), the detailed pathogenesis of glaucoma formation and development is still poorly understood, and one of the major acceptable and effective approach to treat glaucoma patients currently were methods to prevent glaucoma from progressing further.

Different cell types within the retina were theorized to participate in different parts of the pathways that leads to glaucoma development, but the precise roles of these cell types and how they affect each other has yet to be confirmed.

This study utilizes a cutting-edge technology, single cell RNA-sequencing, to analyse glaucoma mouse model tissues. This experimental method allows us to obtain and analyse the transcriptomes of single cells, providing a much higher definition in comparison to traditional methods of bulk RNA-seq, which averages expression profiles of all cells within the tissue and omits the heterogeneity of the retina. With the ability to inspect transcriptomic changes within different cell types when comparing between normal and glaucoma conditions, specific pathways and mechanisms activated can be derived from the ontological analysis data. This study applies this technique to two different glaucoma models, further extending the coverage of understanding among different types of glaucoma. The data obtained from this study lays the foundation of single cell glaucoma research in the Ohnuma Lab, which strengthens the international collaboration between UCL and University of Yamanashi on the studies about *ABCA1* in glaucoma development, as well as providing Santen Pharmaceutical Co., Ltd. with valuable insight about the roles of different cell types in the pathogenesis of glaucoma.

The results of this study proposed neuroinflammation and senescence as the potential mechanisms which leads to retinal ganglion cells degeneration, and the participation of different cell types such as glial cells and immune cells. Data generated and analysed from this analytical study could aid in future studies in developing relevant therapeutical targeting of specific cell types and pathways. Further investigations and developments could potentially translate to preventative measures or treatments for glaucoma, impacting the quality of life of glaucoma patients worldwide.

Acknowledgements

I would like to extend my gratitude to my primary supervisor, Prof. Shin-ichi Ohnuma, for providing me with the opportunity to undertake my PhD in his lab, and his support, teaching, and guidance throughout this project. I am thankful for my funding sponsor, Santen Pharmaceutical Co., Ltd., including Dr. Najam Sharif and Dr. Takashi Ota, for financing this PhD project and providing industrial support throughout the process.

I would like to give special thanks to members of the Ohnuma laboratory, particularly my fellow PhD candidate Mr. Jack Hopkins, and technician Dr. Joanna Nightingale, where their assistance and academic guidance have been extremely crucial. I am also very grateful for our collaborators Dr. Schuichi Koizumi and Dr. Youichi Shinozaki from the University of Yamanashi, for providing the knockout mice, sharing their unpublished research findings and contribution to the collaboration in interpreting data. Thank you to my secondary supervisor Prof. Astrid Limb and Dr. Nikolas Pontikos, for their guidance on my project and thesis during the upgrade assessment. I am thankful for the lab managers Dr. Gopi Damodaran and Dr. Naheed Kanuga for their administrative support, and the supporting staff at the Blizard Institute Genomics Centre: Dr Charles Mein, Ms Eva Wozniak and Mr. Theodoros Xenakis for their assistance for the 10X Genomics single cell experiment. Special thanks to Dr. Karen Eastlake for her guidance on retinal tissue handling.

Finally, I could not have pursued this PhD without the support from my family and friends. I am forever indebted to the unconditional love and care from my father and sister, who were always there for me. Thank you to Sergi, Francesco, Mario and Florian who continued to inspire and motivate me even after our undergraduate journey. I am also thankful for my friends outside of the academia, namely Sunny, Chan, Sixtine and Susan, for the support, encouragement, and joy they've brought me throughout my years in UCL.

Table of Contents

| | |
|----------------------------------------------------------------|----|
| Declaration..... | 2 |
| Abstract..... | 3 |
| Impact Statement | 5 |
| Acknowledgements..... | 7 |
| List of Figures | 12 |
| List of Tables | 17 |
| Abbreviations | 19 |
| Chapter 1. General Introduction..... | 21 |
| 1.1 Preface | 21 |
| 1.2 Glaucoma | 21 |
| 1.2.1 General information | 21 |
| 1.2.2 Types of Glaucoma | 22 |
| 1.2.3 Genetic risk factors of glaucoma | 24 |
| 1.2.4 GWAS identification of potential genes association | 25 |
| 1.2.5 ABCA1 Mouse Model..... | 26 |
| 1.2.6 P2Y6-/- model..... | 31 |
| 1.3 RNA-Sequencing technology | 33 |
| 1.3.1 RNA-sequencing | 33 |
| 1.3.2 Single cell RNA-sequencing..... | 37 |
| 1.3.2.1 Overview | 37 |
| 1.3.2.2 Single Cell isolation techniques | 38 |
| 1.3.2.3 Single cell data analysis | 41 |
| 1.4 Retinal Cell types | 45 |
| Chapter 2: Materials and Methods..... | 48 |
| 2.1 Preface | 48 |
| 2.2 Mouse Methods | 48 |
| 2.2.1 Animal housing and treatment..... | 48 |
| 2.2.2 Eye and Retina dissection | 48 |
| 2.2.3 Retina Dissociation | 49 |
| 2.2.3.1 ABCA1 samples..... | 49 |
| 2.2.3.2 P2Y6 samples..... | 49 |
| 2.2.4 10X Single Cell Chromium system | 49 |

| | |
|-------------------------------------------------------------|------------|
| 2.2.5 Genotyping..... | 50 |
| 2.3 Data analysis methods..... | 51 |
| 2.3.1 Cell ranger | 51 |
| 2.3.2 Seurat | 52 |
| 2.3.2.1 Individual samples of ABCA1 | 52 |
| 2.3.2.2 ABCA1 sample integration..... | 53 |
| 2.3.2.3 ABCA1 sub-clustering | 53 |
| 2.3.2.4 P2Y6 samples pre-processing: SoupX | 54 |
| 2.3.2.5 P2Y6 samples integration | 54 |
| 2.3.2.6 P2Y6 samples subclustering | 55 |
| 2.3.3 Ontological Analysis..... | 55 |
| Chapter 3: ABCA1-/- Experiment | 56 |
| 3.1 Preface | 56 |
| 3.2 Tissue sample preparation | 56 |
| 3.2.1 Genotyping..... | 56 |
| 3.2.2 Tissue dissociation..... | 57 |
| 3.3 Data Analysis | 60 |
| 3.3.1 10X Genomics CellRanger output..... | 60 |
| 3.3.2 Seurat Individual samples..... | 63 |
| 3.3.3 Samples Integration..... | 74 |
| 3.3.4 WT vs KO Ontological analysis..... | 82 |
| 3.3.5 Müller Glia and Astrocyte subclusters | 87 |
| 3.3.5.1 Identifying subclusters..... | 87 |
| 3.3.6 Retinal Ganglion Cells subclusters | 95 |
| 3.3.6.1 Identifying subcluster | 95 |
| 3.3.6.2 RGC-subtype ontological analysis..... | 97 |
| 3.3.7 Immune cells subcluster | 100 |
| 3.4 Discussion..... | 102 |
| 3.4.1 Choice of cell number per samples | 102 |
| 3.4.2 Challenges in identifying cell clusters | 104 |
| 3.4.3 Neuroinflammation induced RGC loss in KO retina | 106 |
| 3.5 Conclusion | 107 |
| Chapter 4: P2Y6-/- Experiment | 108 |

| | |
|----------------------------------------------------------------------------------------|-----|
| 4.1 Preface | 108 |
| 4.2 Tissue Preparation | 108 |
| 4.3 Data Analysis | 109 |
| 4.3.1 10X Genomics CellRanger output | 109 |
| 4.3.2 SoupX – Contamination estimation and removal | 110 |
| 4.3.2.1 Unfiltered samples | 111 |
| 4.3.2.2 Contamination profile and fraction estimation | 111 |
| 4.3.2.3 Contamination removal | 117 |
| 4.3.3 Seurat integration | 120 |
| 4.3.3.1 Pre-filtering | 120 |
| 4.3.3.2 All samples Integration | 120 |
| 4.3.3.3 Subclustering of RGCs | 128 |
| 4.3.3.4 Subclustering of Glial cells | 130 |
| 4.4 Ontological Analysis | 134 |
| 4.5 Discussion | 138 |
| 4.5.1 Sample quality of single cell suspension | 138 |
| 4.5.2 Potential subtypes of MGs | 139 |
| 4.6 Conclusion | 140 |
| Chapter 5: Discussion | 141 |
| 5.1 Preface | 141 |
| 5.2 Key findings | 141 |
| 5.2.1 Chapter 2: ABCA1 ^{-/-} glaucoma mouse model | 141 |
| 5.2.2 Chapter 3: P2Y6 ^{-/-} glaucoma mouse model | 141 |
| 5.3 The power of single cell analysis | 142 |
| 5.4 Pathways activation in ABCA1 ^{-/-} induced glaucoma-like phenotype | 144 |
| 5.5 Limitations | 148 |
| 5.5.1 Importance of sample preparation | 148 |
| 5.5.2 Rod photoreceptor filtering | 149 |
| 5.5.3 Gene dropout and determination of cell types | 150 |
| 5.6 Future work | 151 |
| 5.6.1 Comparison between ABCA1 ^{-/-} and P2Y6 ^{-/-} model data | 151 |
| 5.6.2 Validation of cell type and expression markers | 152 |
| 5.6.3 Identification and application of potential drug targets | 153 |

| | |
|--------------------------------------------------------------------------|------------|
| 5.6.4 Inclusion of other retinal structures | 154 |
| 5.6.5 Advanced single cell analysis techniques | 155 |
| 5.7 Conclusion | 155 |
| Chapter X: Appendices | 157 |
| X.1 R code for Individual ABCA1-/- experiment samples: WT1 example | 157 |
| X.2 R code of ABCA1-/- experiment samples integration | 159 |
| X.3 R code of SoupX analysis for P2Y6-/- experiment data..... | 162 |
| X.4 R code of P2Y6-/- experiment data integration..... | 164 |
| Bibliography | 167 |

List of Figures

Chapter 1

Fig 1.2.2 Anatomy of the eye and aqueous humor regulation pathways.

Fig 1.2.3 Estimated prevalence of POAG with age for men and women combined by ethnicity.

Fig 1.2.5.1 ABCA1^{-/-} retina staining data.

Fig 1.2.5.2 ABCA1 expression staining pattern.

Fig 1.2.5.3 Astrocyte-specific ABCA1^{-/-} retina data.

Fig 1.2.6 P2Y6^{-/-} retina data.

Fig 1.3.1 Overview of RNA-sequencing pipeline

Fig 1.3.2.1 Overview of single cell RNA-sequencing experiment pipeline.

Fig 1.3.2.2.1 Moore's law scaling of single cell RNA-sequencing experiment development.

Fig 1.3.2.2.2 10X Genomics single cell RNA-sequencing workflow.

Fig 1.3.2.3 Overview of single cell RNA-sequencing data analysis pipeline.

Fig 1.4 Schematic representation of the major retinal cell types and their organization in the retina.

Chapter 3

Fig 3.1 PCR genotype results of mouse retinal samples.

Fig 3.2.2.1 Fluorescent cell counter images of dissociated wild type mouse retinal cells.

Fig 3.2.2.2 Fluorescent cell counter image of dissociated ABCA1^{-/-} mice retinal cells.

Fig 3.3.1 ABCA1^{-/-} experiment samples sequencing data statistics.

Fig 3.3.2.1 Example of Elbow plot.

Fig 3.3.2.2: WT1 tSNE projection.

Fig 3.3.2.3: WT2 tSNE projection.

Fig 3.3.2.4: WT3 tSNE projection.

Fig 3.3.2.5: KO1 tSNE projection.

Fig 3.3.2.6: KO2 tSNE projection.

Fig 3.3.2.7: KO3 tSNE projection.

Fig 3.3.2.8: KO4 tSNE projection.

Fig 3.3.2.9: KO5 tSNE projection.

Fig 3.3.2.10: Rod cell markers expression in ABCA1-/- experiment samples.

Fig 3.3.2.11: WT1 non-rod tSNE plot.

Fig 3.3.2.12: WT2 non-rod tSNE plot.

Fig 3.3.2.13: WT3 non-rod tSNE plot.

Fig 3.3.2.14: KO1 non-rod tSNE plot.

Fig 3.3.2.15: KO2 non-rod tSNE plot.

Fig 3.3.2.16: KO3 non-rod tSNE plot.

Fig 3.3.2.17: KO4 non-rod tSNE plot.

Fig 3.3.2.18: KO5 non-rod tSNE plot.

Fig 3.3.3.1: UMAP projection of all cells integrated from ABCA1-/- experiment samples.

Fig 3.3.3.2 Elbow Plot from PCA of the integrated dataset of ABCA1-/- experiment.

Fig 3.3.3.3 UMAP projection of clustered and labelled cells from the integration of ABCA1-/- experiment.

Fig 3.3.3.4 Dot Plot of the integrated samples.

Fig 3.3.3.5 Rod cells and cone cells markers expression.

Fig 3.3.3.6 Müller Glia cell/ astrocyte and microglia cell markers expression.

Fig 3.3.3.7 Bipolar cell and RGC markers expression.

Fig 3.3.3.8 RPE and endothelial cell markers expression.

Fig 3.3.4.1 Neuroinflammation Signalling pathway diagram from IPA.

Fig 3.3.5.1.1 tSNE plot of the Müller glia/ astrocytes subcluster.

Fig 3.3.5.1.2 Müller Glia markers in the MG/A subcluster.

Fig 3.3.5.1.3 Astrocytes markers in the MG/A subcluster tSNE plot.

Fig 3.3.5.1.4 Violin Plot of Müller glia markers (upper row) and Astrocyte markers (bottom row) expression in MG/A subclusters.

Fig 3.3.6.1.1 tSNE subclustering of mixed group containing RGCs.

Fig 3.3.6.1.2 RGC specific markers expression in tSNE subclusters.

Fig 3.3.6.1.3 Violin Plot of RGC specific markers expression in tSNE subclusters.

Fig 3.3.6.2.1 RGC related tsne plots and data

Fig 3.3.6.2.2 KEGG analysis of genes expressed in RGC-1.

Fig 3.3.7.1 tSNE plot and data of immune cell cluster.

Fig 3.3.7.2 Expression levels of P2ry12 and C1qa in Immune Cell cluster.

Fig 3.4.1 Number of differentially expressed genes detected in different cell types.

Fig 3.4.2 Number of cells in different cell types.

Chapter 4

Fig 4.2 Example images from Trypan blue assays of dissociated samples for P2Y6^{-/-} experiment.

Fig 4.3.1 Statistical data of P2Y6^{-/-} samples.

Fig 4.3.2.1.1 UMAP projection of WT samples and Rho expression, likelihood of true Rho expression estimated by SoupX and correction level of Rho in WT samples.

Fig 4.3.2.1.2 UMAP projection of KO samples and Rho expression, likelihood of true Rho expression estimated by SoupX and correction level of Rho in KO samples.

Fig 4.3.2.2 Estimation plot produced by autoEstCont function to predict likelihood of contamination fraction.

Fig 4.3.2.3.1 UMAP projection of corrected WT samples and Rho expression.

Fig 4.3.2.3.2 UMAP projection of corrected KO samples and Rho expression.

Fig 4.3.3.2.1 UMAP plot of integrated cells from P2Y6^{-/-} experiment data.

Fig 4.3.3.2.2 Elbow Plot from PCA of the integrated dataset of P2Y6^{-/-} experiment.

Fig 4.3.3.2.3 UMAP plot of integrated cells from P2Y6^{-/-} experiment data.

Fig 4.3.3.2.4 Dot plot of the integrated P2Y6^{-/-} experiment samples.

Fig 4.3.3.2.5 Rod and cone cell markers expression.

Fig 4.3.3.2.6 Müller glia and RGC markers expression.

Fig 4.3.3.2.7 Bipolar cell, Immune cells, RPE and endothelium cell markers expression.

Fig 4.3.3.3.1 RGC cell marker expression plot of integrated dataset.

Fig 4.3.3.3.2 UMAP Subclustering of cells from cluster 11 & 18.

Fig 4.3.3.3.3 RGC cell marker expression in subclustering plot and subcluster split by origin.

Fig 4.3.3.4.1 UMAP subclustering of cells from cluster 5 & 10.

Fig 4.3.3.4.2 Müller glia subcluster UMAP split by cluster origin.

Fig 4.3.3.4.3 Müller glia cell marker expression plot and dot plot.

Chapter 5

Fig 5.6.2 tSNE projection of the RGC atlas data from Tran et.al. study with integrated cells from ABCA1^{-/-} samples.

Fig 5.6.3 Schematic diagram illustrating the trabecular meshwork aqueous humor production and outflow pathway.

List of Tables

Chapter 2

Table 2.2.5.1 PCR primer sequences

Table 2.2.5.2 PCR reaction mixture

Table 2.2.5.3 PCR thermocycle

Chapter 3

Table 3.2.2 Loading conditions of dissociated retinal cells

Table 3.3.1 CellRanger sequencing output summary of ABCA1^{-/-} experiment samples

Table 3.3.2.1 Retinal Cell Markers used to identify and label subtypes

Table 3.3.3.1 Cell number counts for different cell types

Table 3.3.3.2 Cell counts in percentage for different cell types.

Table 3.3.4.1 Top 50 most significantly affected Canonical Pathways in differential expression between all cells.

Table 3.3.4.2 Neuroinflammation related pathways activated in KO vs WT all cells differential expression.

Table 3.3.4.3 Senescence related pathways activated in KO vs WT all cells differential expression.

Table 3.3.4.4 SASP related pathways activated in KO vs WT all cells differential expression.

Table 3.3.5.2.1 Top 50 most significantly affected Canonical Pathways in differential expression between all Müller glia.

Table 3.3.5.2.2 Top 50 most significantly affected Canonical Pathways in differential expression between all astrocytes

Table 3.3.5.2.3 Neuroinflammation related canonical pathways affected in Müller glia and astrocytes

Table 3.3.5.2.4 Senescence related canonical pathways affected in Müller glia and astrocytes

Table 3.3.6.2.1 Disease and Functions pathway activated in RGC-1 cells of KO samples.

Chapter 4

Table 4.3.1 CellRanger sequencing output summary of P2Y6^{-/-} experiment samples.

Table 4.3.3.1 Filtering values used on SoupX-corrected samples.

Table 4.3.3.2.1 Cell counts of different cell types from P2Y6^{-/-} experiment samples.

Table 4.3.3.2.2 Cell count distribution of different cell types from P2Y6^{-/-} experiment samples.

Table 4.4.1 Canonical pathways activation related to neuroinflammation.

Table 4.4.2 Canonical pathways activation related to senescence.

Table 4.4.3 Canonical pathways activation related to angiogenesis and hypoxia.

Table 4.4.4 Canonical pathways activation related to fibrosis.

Abbreviations

| | |
|-------|--------------------------------------------------|
| ABCA1 | ATP-binding cassette transporter 1 |
| AMD | Age-related macular degeneration |
| ATP | Adenosine triphosphate |
| BCL | Binary base call |
| CCA | Canonical correlation analysis |
| cDNA | complementary DNA |
| CERP | Cholesterol efflux regulatory protein |
| cKO | Conditional knockout |
| FACS | Fluorescence-activated cell sorting |
| GEM | Gel Beads-in-Emulsions |
| GWAS | Genome-wide association studies |
| HDL | High-density lipoprotein |
| IOP | Intraocular pressure |
| IPA | Ingenuity Pathway Analysis |
| KEGG | Kyoto Encyclopedia of Genes and Genomes analysis |
| KO | Knockout |
| MACS | Magnetic-activated cell sorting |
| mfERG | Multifocal electroretinograms |
| MG | Müller glia |
| MO | Months old |
| mRNA | Messenger RNA |
| NGS | Next-generation sequencing |

| | |
|---------|-----------------------------------------------|
| NTG | Normal tension glaucoma |
| PACG | Primary angle-closure glaucoma |
| PC | Principal Components |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| POAG | Primary open-angle glaucoma |
| qPCR | Quantitative polymerase chain reaction |
| RGC | Retinal ganglion cell |
| RNA-seq | RNA sequencing |
| RPCA | Reciprocal PCA |
| SACG | Secondary angle-closure glaucoma |
| SASP | Senescence-associated secretory phenotype |
| SNP | Single nucleotide polymorphisms |
| SOAG | Secondary open-angle glaucoma |
| tSNE | T-distributed stochastic neighbor embedding |
| TUNEL | TdT-mediated dUTP nick end labelling |
| UDP | Uridine diphosphate |
| UMAP | Uniform manifold approximation and projection |
| UMI | Unique molecular identifier |
| WGS | Whole genome sequencing |
| WT | Wild type |

Chapter 1. General Introduction

1.1 Preface

The premise of this PhD thesis is to understand the pathogenesis of glaucoma through the use of single cell RNA-sequencing analysis. This chapter introduces the current studies and understanding about glaucoma, the different types of glaucoma and their characteristics, as well as the risk factors associated to the disease with heavy emphasis on genetic factors. Experimental data from two gene knockout mouse models, extensively studied by our Japanese collaborators, were then discussed to suggest their credibility as animal models of different types of glaucoma. Experimental methods of the RNA sequencing technology are then discussed and demonstrated, particularly with the new advancement in single cell RNA sequencing advancement. It concludes with the explanation of the vast variation of cell types found in the retina and establish the project goals of utilizing single cell analysis to understand cell type specific changes in tissue samples extracted from glaucoma animal models.

1.2 Glaucoma

1.2.1 General information

Glaucoma is characterized as a group of eye diseases where the progressive damage of neurons, particularly retinal ganglion cells (RGCs), lead to vision loss. This disease affects more than 70 million people worldwide, with approximately 10% being bilaterally blind (Quigley & Broman, 2006), making it the leading cause of irreversible blindness in the world. Studies based on population-wide surveys suggest less than 50% of patients are aware they have developed glaucoma (Hennis et al., 2007; Leite et al., 2011), as the disease can remain asymptomatic for a long time until it progresses into advanced stages. Thus, the potential population that can be affected with the disease may be underestimated. However, the detailed pathogenesis of glaucoma formation and development is still poorly understood, hence the major acceptable and effective approach to treat glaucoma patients were methods mostly to prevent glaucoma from progressing further (Schuster et al., 2020).

1.2.2 Types of Glaucoma

There are many different types of glaucoma (Casson et al., 2012), characterized by multitudes of factors. The most common subgroups can be defined as “open-angle” or “closed-angle” glaucoma, describing the relative position between ocular lens and the trabecular meshwork, which formats the fluid passage for aqueous humor outflow (Fig.1.2.2). Open angle glaucoma could also be subcategorized based on other symptoms, such as the level of Intraocular pressure (IOP) or causes of disease onset.

With reference to Fig1.2.2A, IOP describes the fluid pressure built within the eye, created by the balance between secretion flow of aqueous humor from the ciliary body and its drainage through the trabecular meshwork and uveoscleral drainage. Although elevated IOP is a major risk factor for most types of glaucoma, there are also subtypes of glaucoma where such elevation is insignificant even though neuronal damage had been observed.

Secondary glaucoma describes glaucoma that occurs after certain clinical trauma or developed as a side effect from other underlying medical conditions (such as exfoliation syndrome related glaucoma (XFG)), whereas Primary glaucoma develops without a clear cause or origin, thus remain as a major focus for research studies. The most common type of glaucoma is Primary Open angle glaucoma (POAG). Over 80% of glaucoma patients from the United States suffer from POAG (Primary Open-Angle Glaucoma), with a predicted estimation of 3 million people being affected. There are about 480,000 patients diagnosed with chronic POAG in England, with potentially over 50% more patients with POAG that have not been diagnosed or are unaware. Patients of POAG with elevated IOP experience increase in the accumulation of fluid in the anterior chamber, inducing mechanical stress and strain to the rest of the structures of the eye, especially compression and deformation of the optic nerve head. Disruption or damage of the most common neuronal cell type at the optic nerve head, which are the RGCs, results in irreversible obstruction or damage of the axons in the visual pathway (Burgoyne et al., 2005).

Patients with POAG who experiences glaucoma-like neuronal damage without the abnormal elevation of IOP, are classified to suffer from Normal Tension Glaucoma (NTG). NTG are typically associated with consistently having IOP under 21 mmHg, while exhibiting glaucomatous phenotypes such as RGC death and degeneration, as well as vision loss (Mallick et al., 2016). NTG mostly affects around 0.17%-0.67% among the worldwide population (Fujino et al., 2016), but a study in 2004 (Iwase et al., 2004) revealed a significantly higher prevalence of NTG in the Japanese population at 3.6%. Understanding of the cause and trigger for NTG development remains fairly limited, and whether there are any common pathways or mechanisms among different types of glaucoma pathogenesis still remain in question. Chronic patients of all kinds of glaucoma may suffer from irreversible retinal damage and vision loss, impacting their quality of life significantly. Thus, the better understanding of the disease initiation, development and process is crucial to help identify potential therapeutical markers and target molecules.

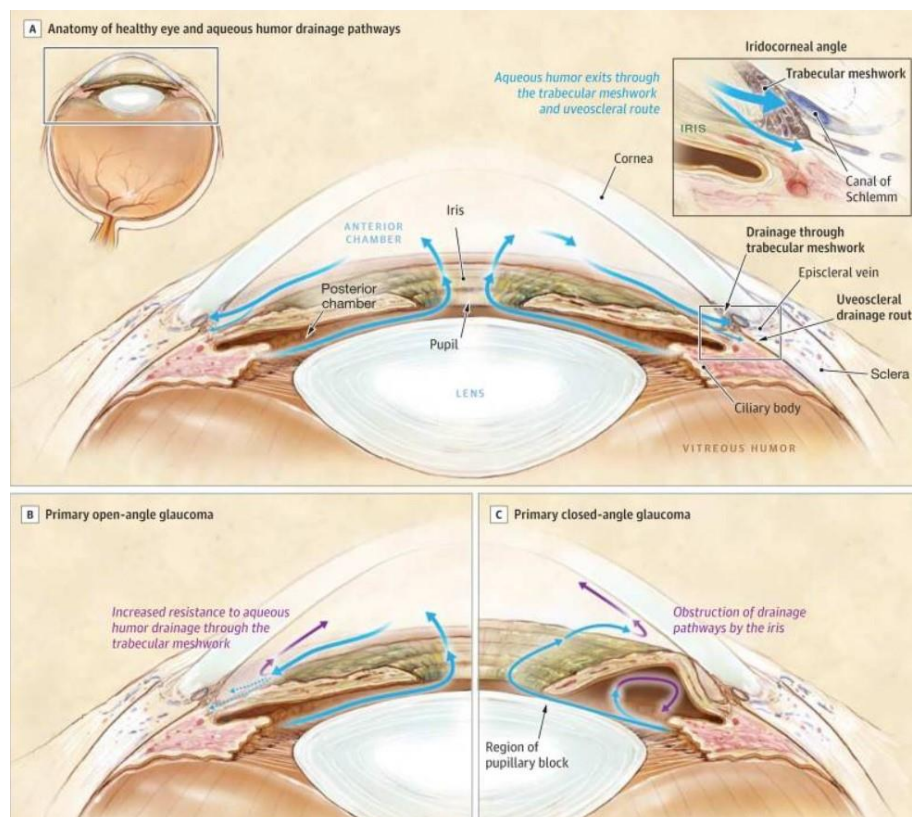


Fig1.2.2 A) Anatomy of healthy eye, and aqueous humor formation & drainage pathways. B) Scenario in Primary open-angle glaucoma. C) Scenario in Primary closed-angle glaucoma (Weinreb et al., 2014)

1.2.3 Genetic risk factors of glaucoma

Glaucoma can also be differentiated based on the age of onset and inheritance form. Some early onset glaucoma such as congenital glaucoma exhibits Mendelian inheritance, while other adult-onset glaucoma such as POAG, PACG are inherited in more complex traits. This suggests glaucoma can affect all ages, but also signifies the importance of genetics in the onset of glaucoma.

Early onset glaucoma which exhibits autosomal recessive traits allows for the identification of causal genes through lineage studies, such as the identification of CYP1B1 and LTBP2 in Congenital glaucoma. (B A Bejjani, 2000; Manir Ali, 2009). However, complex traits of other adult-onset glaucoma can be difficult to decipher, especially the variations among the human population.

A study in 2016 (Kapetanakis et al., 2016) conducted systematic review of published data examining the relationship between POAG and different demographic factors. As suggested in Fig1.2.3, the prevalence of POAG cases increase exponentially with age across all ethnicities, ranging from 0.4-0.8% at the age of 40, 1.4-5.2% at age 60, and 5.5-25.5% at age 90. Other studies have shown people of African descent have a prevalence of POAG 2.8 times higher than that of Europeans, while Asians are 4.2 times more likely to have PACG than North Americans (Tham et al., 2014). Family history studies have discovered around half of all POAG patients have a positive glaucoma family history, with their closest relatives such as parents, siblings and children having a 9-fold increased risk of developing glaucoma in their life (Awadalla et al., 2015).

This significant evidence suggests aging as an important factor for glaucoma development, while ethnicity or family history contributes to the genetic risk for the susceptibility of having the disease. Thus, the studies of genetics and gene expressions in glaucoma patients are exceedingly important in identifying these genetic risk factors among the population, and how aging and development are involved in

triggering the onset of the disease. Genetically modified animal models would hence be effective in studying the disease process and phenomenon.

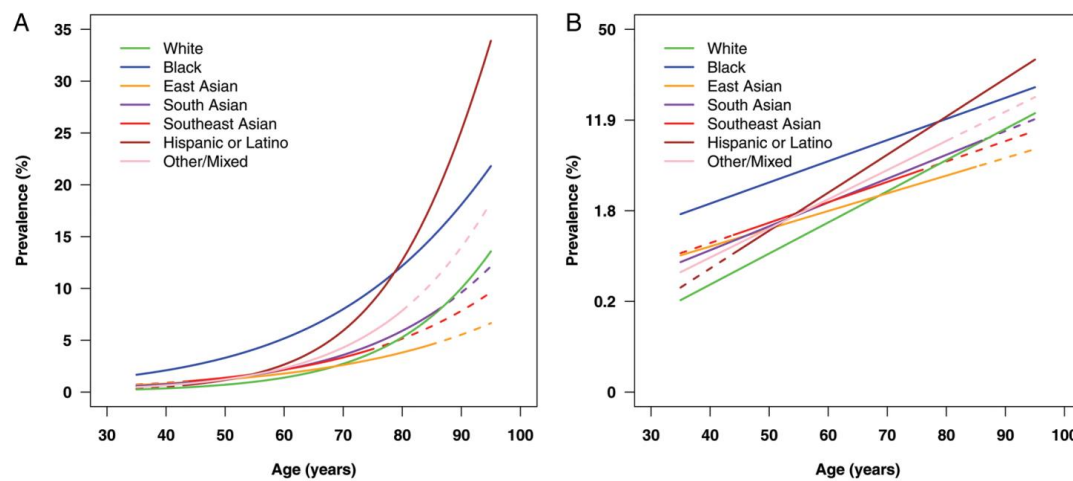


Fig1.2.3 Estimated prevalence (%) of primary open angle glaucoma with age for men and women combined by ethnicity; (A) shows prevalence on the y axis on the normal scale, (B) on the log scale. Coloured lines come from regression models adjusting for age (log-linear relationship), fitted separately for White (green), Black (blue), East Asian (orange), South Asian (purple), Southeast Asian (navy), Hispanic or Latino (brown), and other or mixed ethnicity (pink) groups. Solid lines are given across the age range of available data for each ethnic group. (Kapetanakis et al., 2016)

1.2.4 GWAS identification of potential genes association

To decipher the associations between complex genetic traits with adult-onset glaucoma, scientists have been utilizing Genome-Wide Association Studies (GWAS). GWAS performs observational studies to identify genome-wide variations and associate them with phenotypic traits. Frequencies of genetic variations in similar loci associated with disease phenotypic traits suggests potential roles in disease onset or regulation, thus allowing researchers to try investigate nearby genes in the genome to identify potential genetic markers associated with such diseases. (Tam et al., 2019) The first GWAS study was performed to analyse patients with Age-related Macular Degeneration (AMD), which identified the variants in the region near gene *CFH* being strongly affected (Klein et al., 2005). As the technique and approach of GWAS develops over the last 20 years, multiple studies have been applied for different diseases, including glaucoma.

Multiple large-scale GWAS from independent, international cohorts had identified multiple SNPs on chromosome 9 nearby the gene encoding ATP-binding cassette transporter 1 (*ABCA1*) significantly high correlation with the glaucoma patients (Chen et al., 2014; Gharahkhani et al., 2021). *ABCA1* gene encodes the membrane-associated protein, which functions as an efflux pump of cholesterol and phospholipids. Other studies have shown association between *ABCA1* and other neurodegenerative diseases such as AMD (Chen, 2010) and Alzheimer's disease (Lewandowski et al., 2022; Lupton et al., 2014), while some studies suggested linkage between cholesterol level and risks for glaucoma (Kang et al., 2019; Posch-Pertl et al., 2022). Thus a direct approach to understand the role of *ABCA1* in glaucoma development was needed, which lead to a collaboration study with Prof. Shinozaki and Prof. Koizumi from University of Yamanashi, where they developed and studied an *ABCA1*^{-/-} mouse model.

1.2.5 ABCA1 Mouse Model

ABCA1 knockout (KO) mouse models were created in the DBA1/*lacJ* strain by homologous recombination, resulting in a 910 base pair deletion at exons 17-22 of the gene, as per the methods in Hamon,2000. Glaucoma-like phenotypes were then investigated in *ABCA1*^{-/-} retinal tissues. Firstly, RGCs condition were examined inspecting signals for their marker, *Brn3a*. Fig1.2.5.1a showed no significant reduction of *Brn3a*⁺ signals in 3 months old (mo) KO retinal samples, while Fig1.2.5.1b showed a significant reduction in 12mo samples. TdT-mediated dUTP nick end labelling (TUNEL), which highlights apoptotic cells, were performed as shown in Fig1.2.5.1c. 12mo KO retinal samples showed significant increase in TUNEL⁺ signals, and these signals co-localize with *Brn3a*⁺ cells as shown in Fig1.2.5.1d. The abundance of TUNEL⁺/*Brn3a*⁺ cells show an age-dependent RGC degeneration under the effect of *Abca1* KO, resembling glaucoma phenotype. To determine which type of glaucoma does the *Abca1*^{-/-} mouse resemble, IOP was measured by a rebound tonometer on both 3mo and 12mo KO mice. Fig1.2.5.1e suggested no significant reduction in IOP for both KO samples, thus signifying the RGC damage induced by the deletion of *Abca1* was not caused by IOP elevation. Hence, the *Abca1*^{-/-} mice were considered a mouse model for Normal Tension Glaucoma (NTG).

As the expression of *Abca1* is related to the glaucoma-like phenotype, cells expressing *Abca1* would be of interest to investigate. Fig1.2.5.2a showed *Abca1* expression mostly at the inner surface of the retina, co-localized with *Gfap* signal, a marker for astrocytes. Fig1.2.5.2c also labelled co-expression of *Abca1* and *Gfap* in human retinal tissues. These findings align with other studies looking at gene expression in both mouse and human, consolidating the expression of gene *Abca1* is most highly expressed in astrocytes (Y. Zhang et al., 2014; Y. Zhang & Wallace, 2015). To consolidate the role of astrocytes in *Abca1* mediated glaucoma pathogenesis, researchers created and inspected astrocyte-specific *Abca1* KO mouse models.

Astrocyte-specific conditional ABCA1 KO (cKO) was achieved by crossing ABCA1^{flox/flox} mouse line with GFAP-Cre mouse. ABCA1^{flox/flox} line was created by flanking exon 45-46 of the gene with a loxP site (Timmins et al., 2005). The transgenic construct containing the *GFAP* promoter and cDNA encoding Cre recombinase were present in the GFAP-Cre mouse line (Bajenaru et al., 2002). Offspring of the crossing allows for cells expressing *GFAP*, which are astrocytes in the retina, to express the Cre recombinase which eliminates exon 45-46 of *ABCA1* and result in the inactivation of the protein in those cells. Efficacy of the system was verified by our collaborators.

Similar to non-specific KO mice, 3mo cKO mice showed no changes in the number of *Brn3a*⁺ RGCs (Fig1.2.5.3a) or apoptotic cells (Fig1.2.5.3b). However, 12mo cKO mice showed a moderate (~20%) but significant reduction in RGC number (Fig1.2.5.3c) and a significantly increase of TUNEL⁺ cells (Fig1.2.5.3d). IOP again did not vary significantly between WT and KO for both 3mo and 12mo samples (Fig1.2.5.3e). Multifocal electroretinograms (mfERG) were used to assess visual function of the mice, and the average visual response of 12mo cKO mice were significantly reduced (Fig1.2.5.3f).

These findings confirmed the importance of *Abca1* expression in retinal astrocytes, and that the astrocyte-specific KO of the gene results in glaucoma-like phenotypes with insignificant IOP elevation in mice. Something noteworthy is that although the association between *Abca1* and glaucoma in both mice and

humans has been highly suggested, the role of IOP regulation by *Abca1* in human may be in contradiction. Studies have demonstrated through human GWAS the association between loci near *Abca1* and risks of IOP elevation (Pirro G Hysi, 2014), while Hu et. al provided experimental evidence on the role of IOP regulation by *Abca1* via the caveolin-1/endothelial NO synthase/NO pathway (Chunchun Hu, 2020). This indicates some potential differences between the mode of action of *Abca1* within the mouse model selected and human patients, and although this study regard *Abca1*^{-/-} mice as NTG model, its effect and significance in astrocyte physiology and glaucoma disease onset may be applicable and be expanded to other POAG with elevated IOP in human. Understanding the mechanisms or pathways within astrocytes that involve *Abca1* is thus crucial to understanding how astrocyte was affected, and how the affected astrocytes contribute to apoptosis and degeneration of other cell types in the retina such as RGCs. Thus, studying cellular physiology and gene expression of different cell types within the retinal tissue becomes very useful and important.

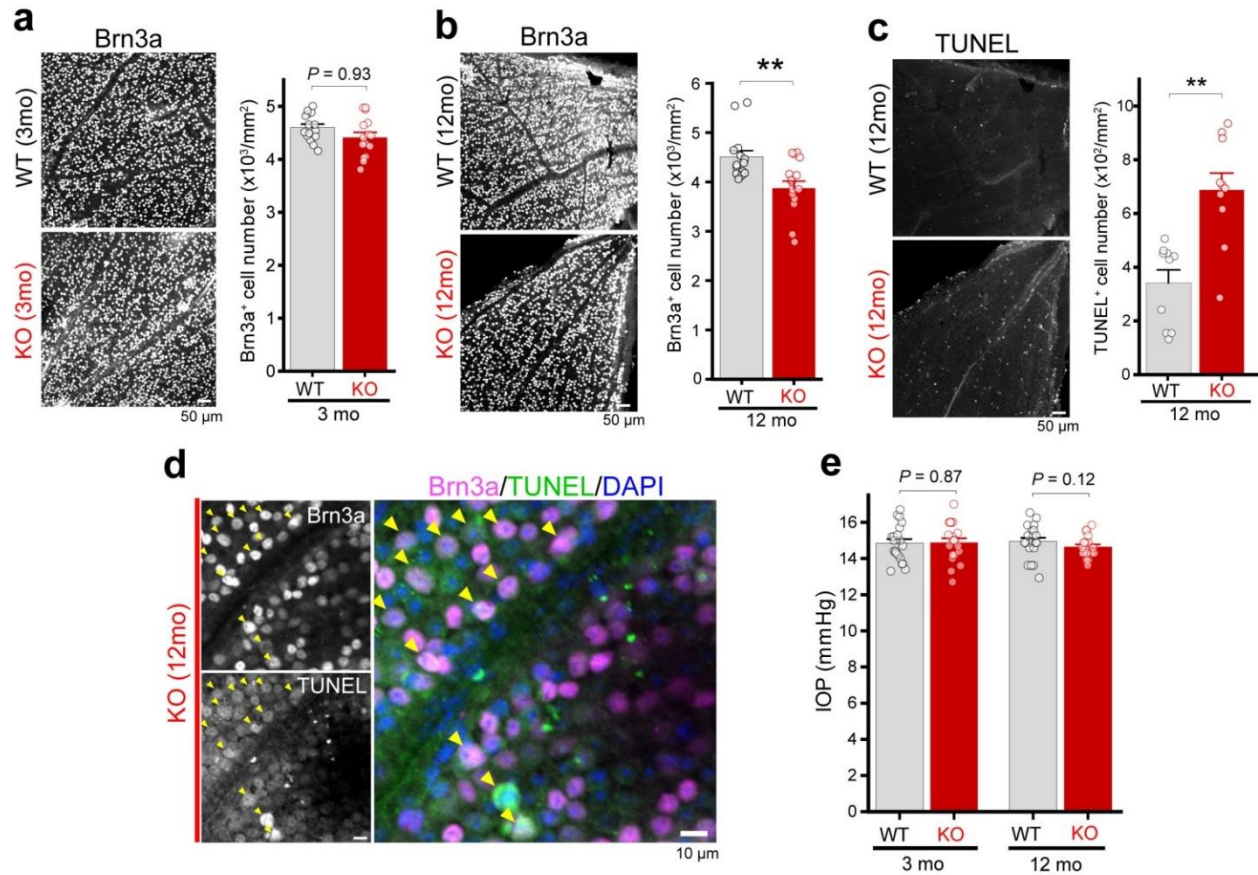


Fig1.2.5.1a: Brn3a signals in retinas from 3 months old WT and KO mice. b: Brn3a signals in retinas from 12 months old WT and KO mice. At 3 months old, the number of Brn3a⁺ cells did not differ between WT and KO, but at 12 months old, Brn3a⁺ cell numbers was reduced. c: TUNEL⁺ cell number was significantly higher in 12 mo KO mice compared with WT mice. d: TUNEL and Brn3a signals were co-localized (arrow); TUNEL⁺/Brn3a⁺ cells were abundant in KO mice. (e) KO mice showed no IOP change.

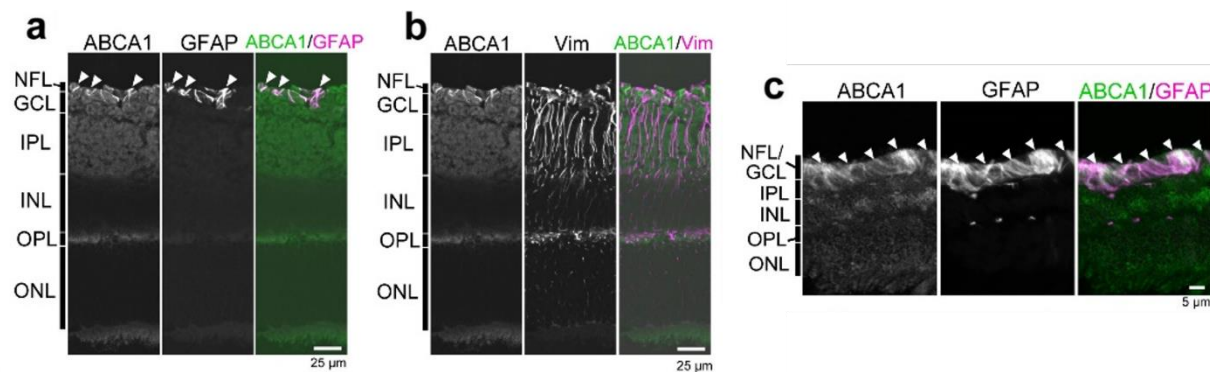


Fig1.2.5.2(a, b) ABCA1 protein expression patterns in retinal slices of 3 mo WT mice. ABCA1 signals in the retina were colocalized with GFAP (arrows) but not with vimentin (Vim), a Müller glia marker. (c) ABCA1 was also expressed in GFAP⁺ astrocytes in the human retina.

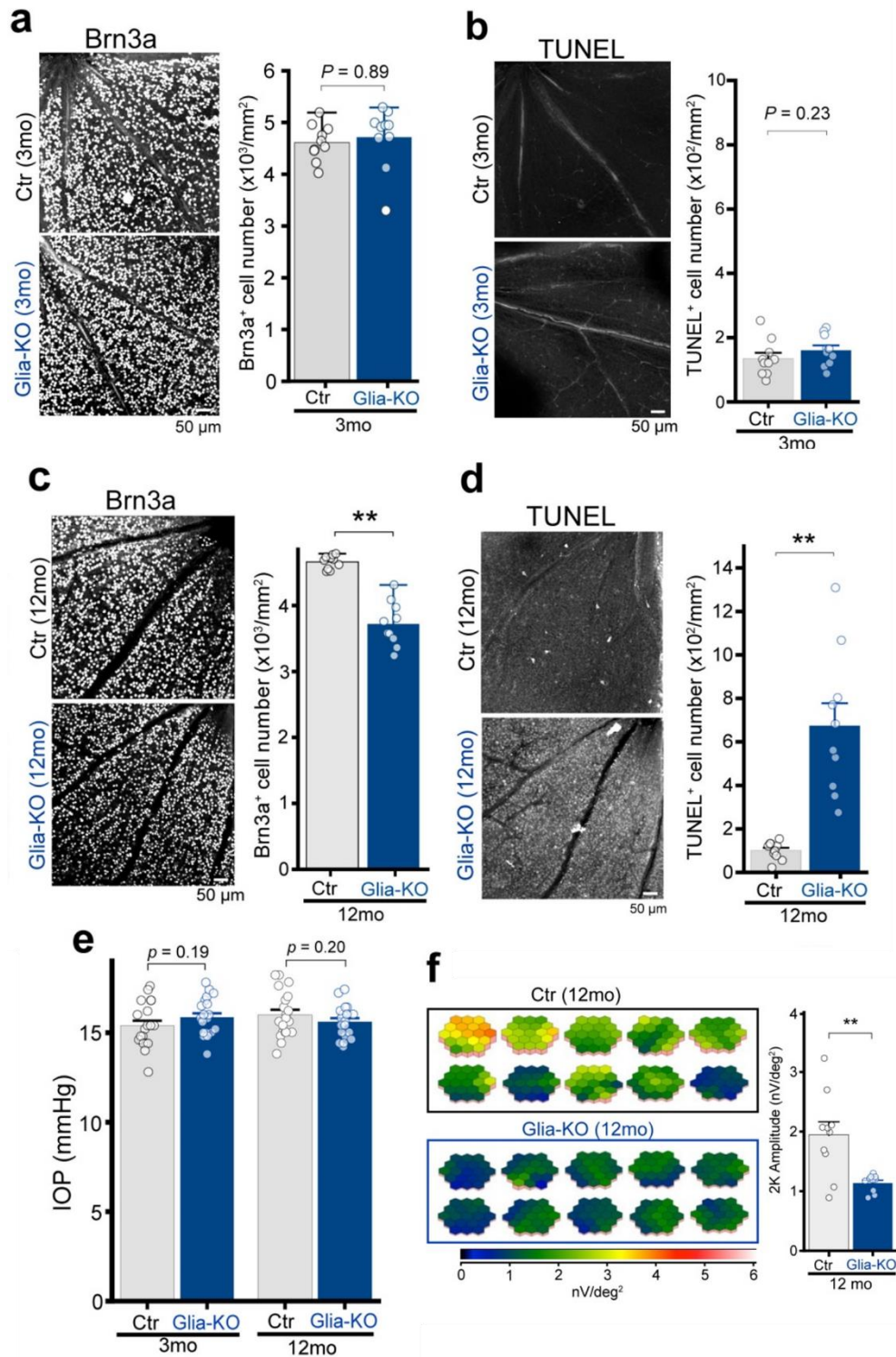


Fig1.2.5.3 At 3 months of age, cKO mice showed (a) no reduction in Brn3a⁺ cells and (b) no increase in TUNEL⁺ cells compared with WT mice. cKO mice at 12 months old showed (c) reduced or (d) increased Brn3a⁺ RGC or apoptotic cell numbers, respectively (e) cKO mice showed no IOP changes compared with age-matched WT mice. (f) Impaired ocular function estimated by multifocal electroretinograms. Left panels: The three-dimensional plots show averaged visual responses of the second-order kernel (2K amplitude). Right: Ocular responses in 12 mo cKO mice were significantly reduced compared with WT mice.

1.2.6 P2Y6^{-/-} model

ABCA1^{-/-} mouse model data suggested strong evidence and similarity as a NTG model for mice, which provides retinal tissue for the study of the pathogenesis of the disease. However, the most common type of glaucoma such as POAG involves stronger IOP changes, and the pathways or molecules involved in glaucoma triggering or development may differ or share similarities. Studying of different animal models of different types of glaucoma could provide the opportunity to understand common shared mechanisms that lead to RGC degeneration, which could be beneficial in identifying therapeutical targets that can benefit in treating all types of glaucoma.

Increase in intraocular pressure is due to the imbalance between aqueous humor production and drainage. Unusual increase in aqueous humor production or decrease in the rate of its draining away from the eye could lead to the building up of fluid, causing the increase in IOP. Nucleotides and its derivatives such as adenosine triphosphate (ATP) or Uridine diphosphate (UDP) were suggested to be involved in the regulation of IOP, especially as elevated levels of ATP were found in glaucoma patients (Li et al., 2011; Markovskaya et al., 2008; X. Zhang et al., 2007). These molecules bind to P2 receptors, which are divided into P2X and P2Y subfamilies, and they are found to be expressed in tissues within the eye such as the cornea, ciliary processes, trabecular meshwork, photoreceptors, and ganglion cells (Pintor et al., 2004). Of these receptors, our collaborators performed extensive studies on the activity of *P2Y6* receptors and its role on aqueous humor production and IOP level manipulation.

Their publication in 2017 (Shinozaki et al., 2017) showed the application of UDP, the endogenous agonist of *P2Y6* receptors, reduced IOP in WT mice by reducing aqueous humor production, similar to the effects of Timolol, which is a commercially available beta-blocker drug used to relieve IOP and treating glaucoma. *P2Y6* knockout mouse (*P2Y6KO*) were then created (Bar et al., 2008) and studied, and the application of UDP did not change the level of aqueous humor production. *P2Y6 KO* mice showed chronic IOP elevation

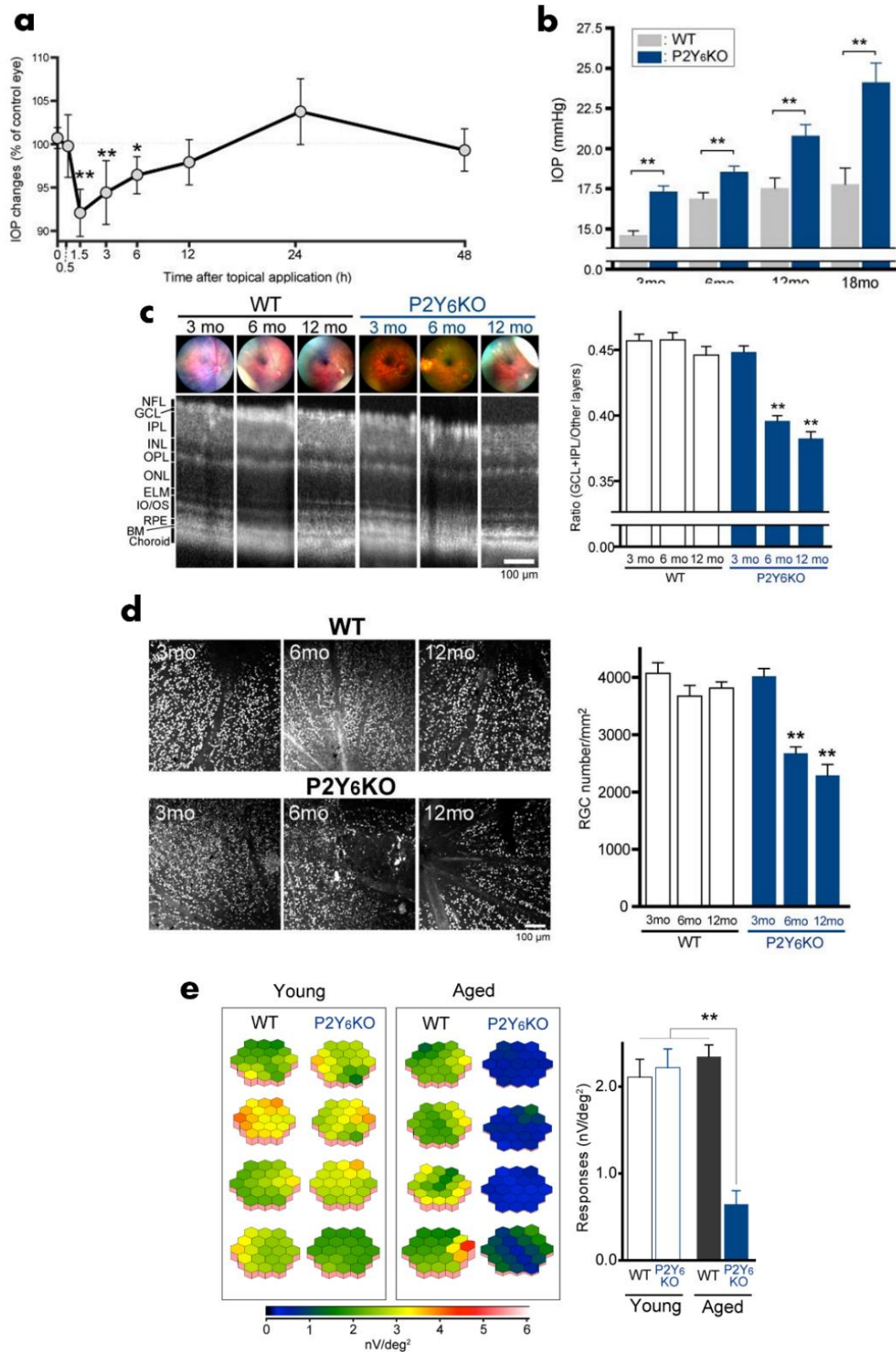


Fig1.2.6a) IOP reduction after topical application of UDP on the retina. b) Chronic IOP increase in aging mice of P2Y6KO mice. c) Retinal thinning, d) RGC degeneration and e) ocular dysfunction observed in aging KO mice. (Shinozaki et al., 2017)

which increases by age, as well as high IOP glaucoma-like phenotypes such as retinal thinning, RGC degeneration and ocular dysfunction (Fig1.2.6). Their research provided solid foundation and evidence on the role of gene *P2Y6* in IOP regulation and aqueous humor production, thus allowing us to use tissues from *P2Y6KO* mouse samples to study the physiological and genetic changes of retinal neurons under the pressure of high IOP.

1.3 RNA-Sequencing technology

1.3.1 RNA-sequencing

To perform a comprehensive study of a disease such as glaucoma, where genetics are a risk factor, research and study of the genetic information available in patients or disease models are crucial. The central dogma of molecular biology suggests the flow of genetic information originates in DNA, which is transcribed into RNA and translated into protein (Crick, 1970). The selected expression of a subset of the DNA genome in the form of transcripts such as messenger RNA (mRNA) encodes different protein within different cell types and tissues, contributing to different functions such as providing structures, facilitate signaling, transporting molecules, aiding in identification and involvement in metabolic reactions. The collected profile of transcript expression within a cell, known as the transcriptome or expression profile, provides insight into its different biological functions and phenotypical behaviour. Thus, the technique or technology of quantifying cellular transcriptome allows researchers to explore biological questions such as disease development on a molecular level.

Early genetics study techniques include northern blots, quantitative polymerase chain reaction (qPCR) and microarray. These methods, however, comes with limitations such as low throughput, prior knowledge in sequence of interest, or limited accuracy in quantifying very lowly or highly expressed genes (Casneuf et al., 2007; Kukurba & Montgomery, 2015; Shendure, 2008). Sequence-based approaches of quantifying

transcriptomes then began to develop, along with the development of high-throughput next-generation sequencing (NGS) technology. RNA sequencing (RNA-seq) utilizes NGS by constructing complementary DNA (cDNA) libraries from reverse transcription of RNA in much high throughput and accuracy. Accompanied by advancement in analytical methods in processing large amount of genetic data, i.e. the field of bioinformatics, RNA-Seq workflows revolutionized the complexity and depth of transcriptomic studies.

A typical overview of the wet lab workflow of RNA-seq is shown in Fig1.3.1a. The most common RNA-seq experiment begins with the isolation of cells or tissues of interest, then extracting RNA materials usually by lysing cell membranes. Depending on the objectives, selection of different kinds of RNA molecules can be performed, although most transcriptomic information of interest originates from mRNAs. Selected RNAs are then converted into cDNA by reverse transcription and ligated with sequencing adaptors to construct the sequencing library. Amplification of the library by polymerase chain reaction (PCR) creates more copies of the transcripts which facilitate sequencing using NGS methods.

Subsequent workflow in RNA-Seq mainly focuses on the application of bioinformatics and computational methods to analyse sequencing data, suggested in Fig1.3.1b. Popular NGS platforms such as Illumina outputs binary base call (BCL) files which are then converted into FASTQ files, containing individual base pair on the reads accompanied with its confidence or quality score. Programs such as FASTQC provides insight into statistics about the sequencing runs per sample, and offer options to perform quality control. Satisfactory sequencing reads were then aligned to respective reference genome of their species origin, namely with algorithms such as STAR (Dobin et al., 2013), TopHat (Trapnell et al., 2009) or Hisat2 (Kim et al., 2019). Mapped reads are then assembled into transcripts, either to the reference genome or using de-novo assembly methods. Quantification of the transcripts were then performed by tools such as Cufflinks (Trapnell et al., 2010), HTSeq (Anders et al., 2013) or featureCounts (Liao et al., 2014), thus creating

estimates of the expression level of different genes within the cell or tissue. After obtaining expression profiles of different conditions, such as cells or tissues from wild-type or diseased model such as knockout samples, differential expression can be determined by comparing differences in the levels of gene expression. Varieties of packages have been developed to implement different statistical methods of comparison, such as DESeq (Love et al., 2014), Cuffdiff2 (Trapnell et al., 2013) or EdgeR (Robinson et al., 2010). Identification of gene expression differences provided by RNA-seq allows researchers to inspect molecular changes under developmental or diseased conditions, and further deduce potential pathways or processes affected by performing gene ontology analysis.

Most generic or common RNA-seq experiments involve tissue samples which contain large number of cells, where all RNA molecules are pooled together for sequencing, thus sometimes referred to as “bulk RNA-seq”. Although bulk RNA-seq provides great insight and benefits into studying the molecular biology of tissue or cell samples as a whole, expression profile obtained from such experiments may only represent the average expression of genes among different cell types within the sample. For tissues which contain a higher heterogeneity, expression patterns of different cell types may not be obtainable from the averaged transcriptome from bulk RNA-seq. In particular, mammalian retina has at least 55 different distinct cell subtypes (Masland, 2001), but rod photoreceptors alone take up over 60% of the whole population (Lukowski et al., 2019; Macosko et al., 2015b; Menon et al., 2019). Average comparison of bulk RNA-seq expression profiles between WT or KO retinal samples will likely be dominated by variations in the major cell types, thus making it difficult to study other cell types with a smaller population such as astrocytes or RGCs. To further empower the ability to study molecular expression using RNA-seq with a higher resolution and individual level, scientists developed the technique of performing single cell RNA-sequencing.

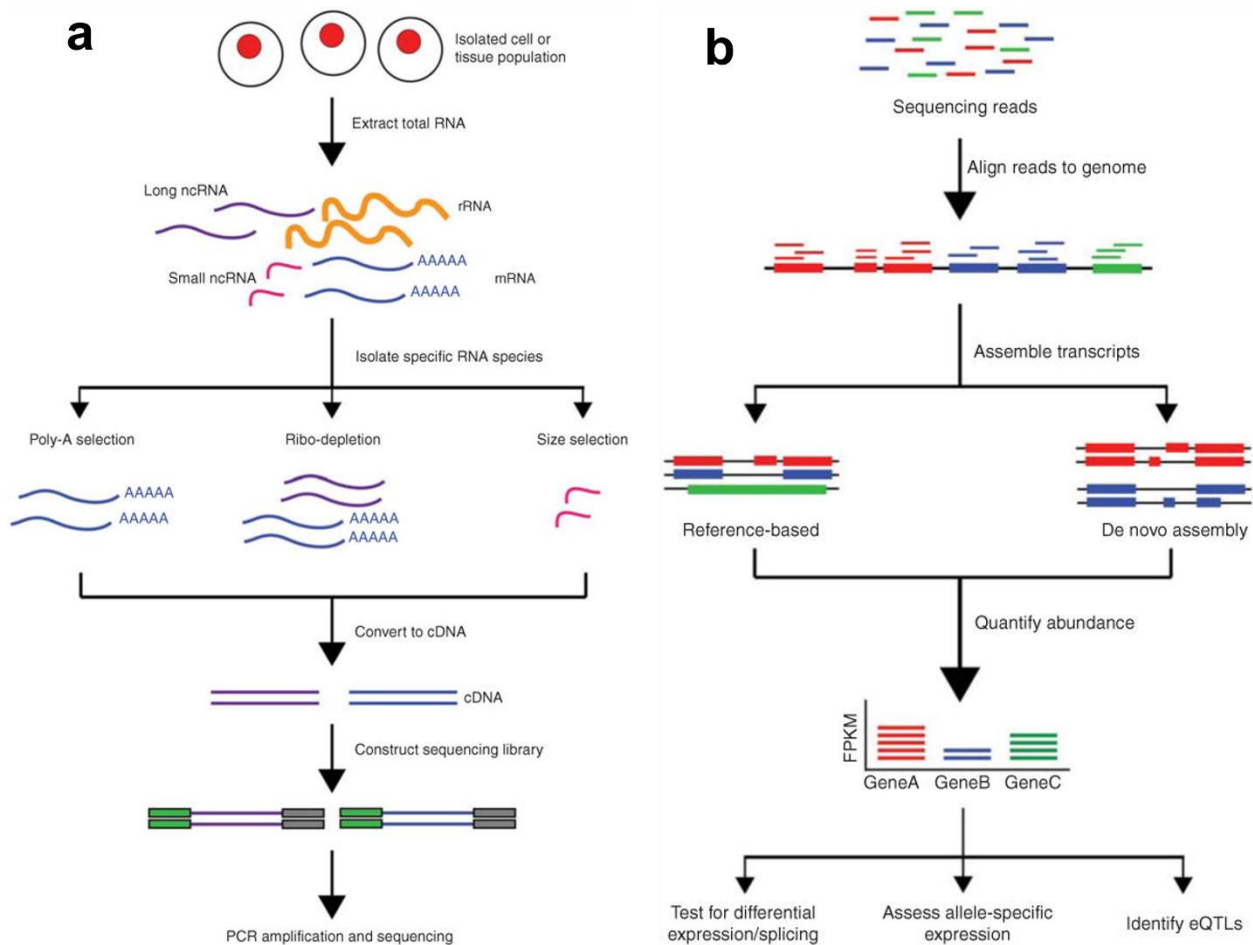


Fig1.3.1a: Overview of RNA-Seq pipeline. First, RNA is extracted from the biological material of choice (e.g., cells, tissues). Second, subsets of RNA molecules are isolated using a specific protocol, such as the poly-A selection protocol to enrich for polyadenylated transcripts or a ribo-depletion protocol to remove ribosomal RNAs. Next, the RNA is converted to complementary DNA (cDNA) by reverse transcription and sequencing adaptors are ligated to the ends of the cDNA fragments. Following amplification by PCR, the RNA-Seq library is ready for sequencing. b: Overview of RNA-Seq data analysis. Following typical RNA-Seq experiments, reads are first aligned to a reference genome. Second, the reads may be assembled into transcripts using reference transcript annotations or de novo assembly approaches. Next, the expression level of each gene is estimated by counting the number of reads that align to each exon or full-length transcript. Downstream analyses with RNA-Seq data include testing for differential expression between samples, detecting allele-specific expression, and identifying expression quantitative trait loci (eQTLs). (Kukurba & Montgomery, 2015)

1.3.2 Single cell RNA-sequencing

1.3.2.1 Overview

The objective of single cell RNA-sequencing (scRNA) was to obtain transcriptomic data of individual cells within a sample, facilitating study of heterogeneous systems such as developmental processes or complex organs or tissues. The first scRNA experiment was originated from a study by Tang et al., 2009, where they obtained the expression profile of a single mouse blastomere. As the technology and science developed over the years both in single cell isolation techniques and sequencing efficiency, more cells can be obtained per sample while cost per cell decreases, making the technology extremely popular. With a larger number and variation of cells being sequenced within each sample, expression distribution among different cell types can be deduced, thus answering new biological questions such as cell-specific changes under disease conditions, identification of new cell types using novel biomarkers, or understanding differentiation pathways and expression markers during early development.

Fig1.3.2.1 shows the 4 key aspects involved in the scRNA workflow. Sample preparation is the first and perhaps most important step, where complex tissue samples are dissected and dissociated. Most scRNA experiments involve dissociation of tissues using enzyme, with the goal to separate cell-cell adhesion while preserving cells as intact individuals. Clumps of cells are ideally filtered out to obtain a single cell suspension of the dissociated tissue, with the options to perform selection for the desired sub-population using techniques such as fluorescence-activated cell sorting (FACS) or magnetic-activated cell sorting (MACS). Single cells are then separated via different cell capturing methods (further discussion in next section), where RNA materials can be obtained individually after cell membrane are lysed. Sequencing library is then performed similarly to bulk RNA-seq, except RNA reads from individual cells are tagged with unique cell barcode. These barcodes allow for the identification of the origin of the reads in the next step during data processing, which facilitate NGS by pooling all RNA reads from different cells in a sample together. Sequenced reads are then processed with bioinformatics pipeline, which first involves

demultiplexing reads using unique cell barcodes to infer their origin. Reads from each cell can then be processed as individual bulk RNA-seq sample, such as mapping to genome, transcript assembling and quantification. Depending on the number of cells and sequencing depth used in different samples, subsequent data analysis involves much more complex methods and algorithms to normalize, quality control, cluster and visualize large amount of sequencing data.

1.3.2.2 Single Cell isolation techniques

Isolation of single individual cells from suspended cell solution is the hallmark of this technology. Experimental techniques of capturing single cells have been evolving exponentially throughout the past decade, and the ingenuity of different single cell capturing methods allows for the exponential increase in the amount cells captured using each protocol, as suggested by the study shown in Fig.1.3.2.2.1. Researchers began isolating single cells manually, then progressed to utilizing microwells, as well as microfluidic platforms, until recently where droplet-based methods became popular. With well-based platforms, cells are mostly isolated by pipetting or laser capturing (FACS) and sorted into microfluidic wells. This method allows researchers to isolate specific subset of cells, either through inspecting microscopic images or selected based on cell surface markers. This process is however fairly tedious, and the number of captured cells remain very low. Microfluidic platform developed by Fluidigm utilizes microfluidic circuits in an integrated chip that captures various single cells in individual reaction chambers and carry out reactions to obtain individual RNA reads. They provide a higher throughput than microwells and the system is highly automated, as well as imaging capability to assess cell conditions within the chamber. However, the chips are relatively expensive due to its intricacy, while the chambers are restricted to handle specific cell size ranges. Samples which include cells of various sizes may face issues such as clogging up the arrays, resulting in a low capture rate.

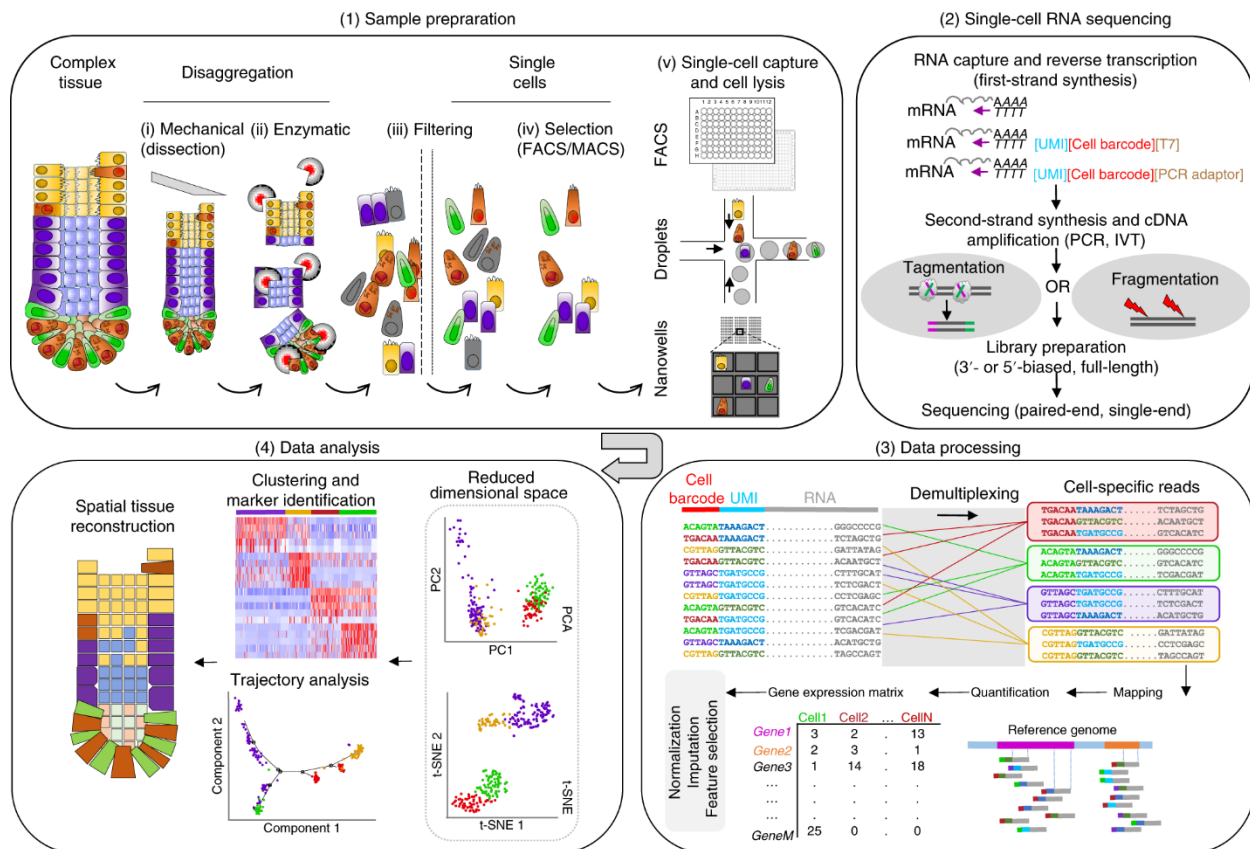


Fig1.3.2.1 The successful design of single-cell transcriptomics experiments includes four major phases: (1) During sample preparation, cells are physically separated into a single-cell solution from which specific cell types can be enriched or excluded (optional). After they have been captured in wells or droplets, single cells are lysed, and the RNA is released for subsequent processing. (2) To convert RNA into sequencing-ready libraries, poly(A)-tailed RNA molecules are captured on poly(T) oligonucleotides that can contain unique molecular identifier (UMI) sequences and single-cell-specific barcodes (5'- and 3' -biased methods). To allow for subsequent amplification of the RNA by PCR or IVT, adaptors or T7 polymerase promoter sequences, respectively, are included in the oligonucleotides. After RT into cDNA and second-strand synthesis (optional), the transcriptome is amplified (PCR or IVT). For conversion into sequencing libraries, the amplicons are fragmented by enzymatic (e.g., tagmentation) or mechanical (e.g., ultrasound) forces. Sequencing adaptors are attached during a final amplification step. Full-length sequencing can be carried out, or 5' or 3' transcript ends can be selected for sequencing using specific amplification primers (optional). For most applications, paired-end sequencing is required. (3) The sequencing reads are demultiplexed on the basis of cell-specific barcodes and mapped to the respective reference genome. UMI sequences are used for the digital counting of RNA molecules and for correction of amplification biases. The resulting gene-expression quantification matrix can subsequently be normalized, and missing values imputed, before informative genes are extracted for the analysis. (4) Dimensional-reduction representations guide the estimation of sample heterogeneity and the data interpretation. Data analysis can then be tailored to the underlying dataset, which allows cells to be clustered into potential cell types and states, or ordered along a predicted trajectory in pseudotime. Eventually, the spatial cellular organization can be reconstructed through the interrogation of marker genes (experimentally) or through marker-guided computational reconstruction (inference). PC, principal component. (Lafzi et al., 2018)

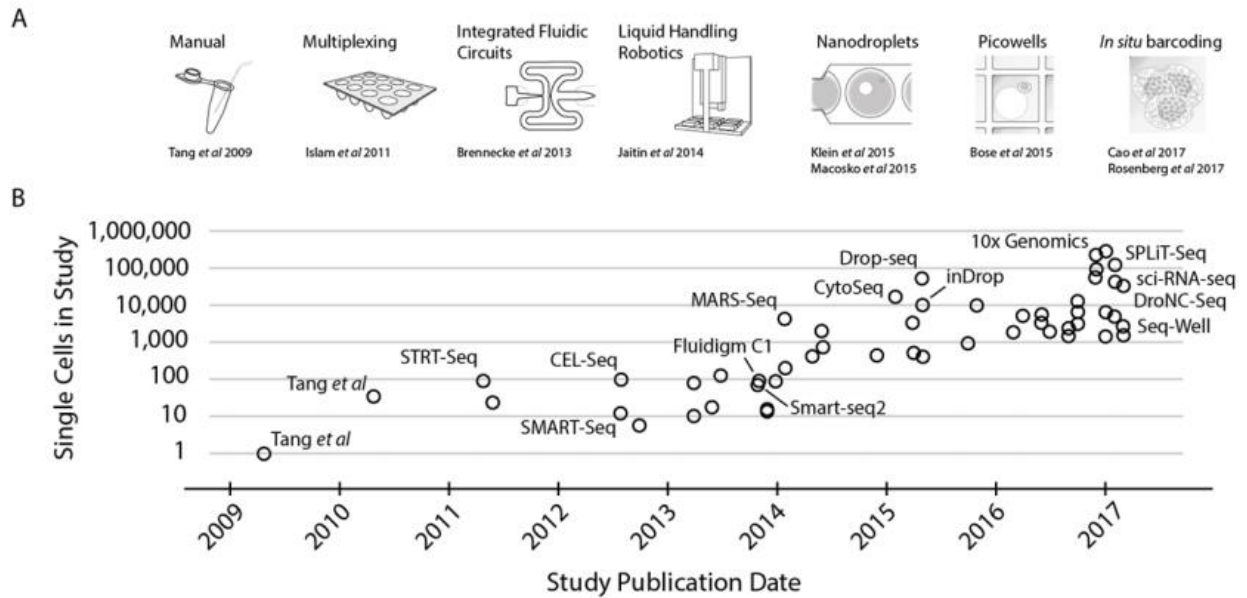


Figure 1.3.2.2.1 Moore's law scaling of scRNA-seq experiments development. (A) Key technologies allowing jumps in experimental scale. (B) Cell numbers reported in representative publications by publication date. (Svensson et al., n.d.)

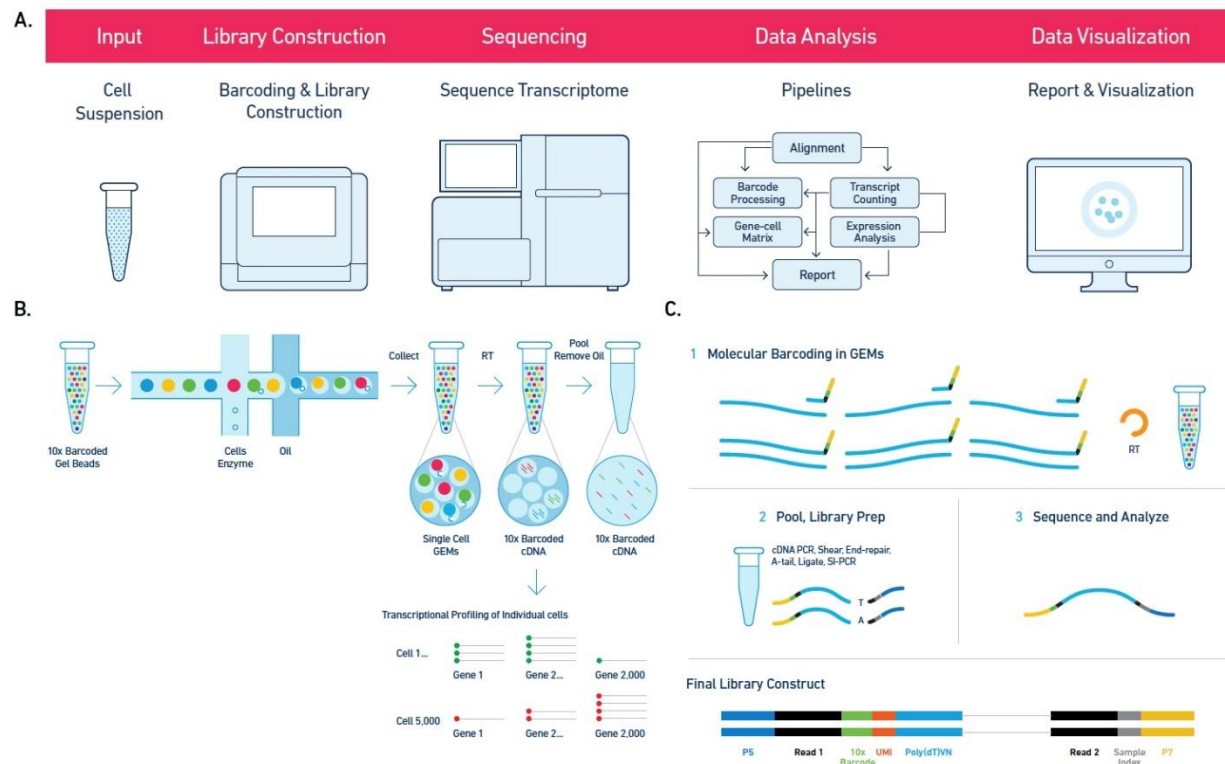


Fig1.3.2.2.2: 10X Genomics workflow B: Sample processing workflow within 10X Chromium system. C: Sequence library preparation workflow. (10X Genomics)

The more advanced, hence popular, droplet-based method allows the capturing of significantly higher number of cells. 10X Genomics achieved so with their “Chromium Single Cell Gene Expression Solution” toolkit (Fig.1.3.2.2.2), where dissociated cells from tissue samples are loaded into the machine and partitioned into nanoliter-scale Gel Beads-in-Emulsions (GEMs) droplet. Each droplet surface is attached with a unique cell barcode with different unique molecular identifier (UMI), along with other enzymes and molecules that facilitate the cell membrane lysis and reverse transcription. Reads from each captured cell will be tagged by the barcode, and the individual sequencing libraries can be created within the cell. Sequencing libraries from different cells can then be pulled together and sequenced in the same lane, high-throughput sequencing can be achieved using NGS methods with cells are lysed within relatively lower cost per cell.

1.3.2.3 Single cell data analysis

As previously mentioned, bulk RNA-seq data are usually processed on a sample level, with analysis pipeline consisting of initial quality control, reference genome alignment, transcript assembly and quantification. Although analysis of scRNA data also involve these processes to obtain transcript details, the complexity and complications in obtaining useful biological information is much higher, due to the sheer number of cells involved in each sample and technical related limitations. Thus the development of bioinformatics toolkits and algorithms are an active field of research, where extra data processing and analysis are involved to fully utilize information provided from scRNA experiments. Fig.1.3.2.3 provides a generic schematic of a scRNA analysis pipeline.

Typical droplet-based scRNA sequencing data analysis first involves demultiplexing reads to identify their origin cell, using the unique barcode sequences they are attached to within each individual droplet. Reads sharing the same barcode can be treated as a small bulk-RNA seq sample, performed with genome alignment, transcript assemble and quantification, until we obtain a massive expression matrix where columns represent individual barcoded droplet, rows contain genes detected, and its element represents

the number of genes detected in each droplet. Different properties due to characteristics and limitations of droplet-based scRNA datasets can be inspected from the expression matrix to perform quality control.

Firstly, defining whether reads from a droplet originate from a single cell is important for constructing the dataset. Although droplet-based scRNA-seq aims at capturing only one cell in one droplet, artifacts may happen where some droplet captures no cell, while some contains two (named doublets) or more cells due to incomplete dissociation. Droplets containing no cell may still capture RNA materials from the background (e.g. RNA from dead cells within the solution). Thus, a prefiltering can be applied based on number of reads or genes detected within each droplet, where significantly higher or lower amount can be discarded as doublets or empty cells respectively. Broken or dying cells usually retain high ratio of mitochondrial RNAs in comparison to cytoplasmic RNAs (Bacher & Kendzierski, 2016), thus removing high mitochondrial RNA content droplets ensures the dataset only contains fresh living cells that were intact at the moment of capture.

Secondly, accurate transcript quantification is much more challenging for scRNA-seq experiments, as the amount of RNA material available per cell is significantly lower than that of a sample in bulk RNA-seq. Effects such as amplification bias (unequal amplification of certain genes during PCR procedures), gene dropout (observation of moderate or high expression of a gene in one cell but not detected completely in another cell), batch effects (technical variability between samples), unwanted biological variations (such as different stages of cell-cycle and genders of cells) or background contamination (inclusion of background RNA molecules from sample preparation solution) result in a much higher level of technical noise in determining the actual transcriptomic profile of different cells. Multitude of approaches are being developed to tackle each level of transcript estimation challenges and attempts to normalize gene expression. For instance, UMIs are unique sequences attached to individual RNA molecules during library construction, such that deduplication can be applied to sequences to remove amplification bias. Approaches such as DESeq2, TMM, SCnorm, SAMstr (Bacher et al., 2017; Katayama et al., 2013; Love et

al., 2014; Robinson & Oshlack, 2010) attempts to perform within-sample and between-sample normalization, while MNN and kBET (Büttner et al., 2018; Haghverdi et al., 2018) corrects batch effects. SoupX (Young & Behjati, 2018) estimates ambient RNA contamination using expression profile from empty droplets, thus correcting cell transcriptome by removing background RNA profile.

The filtered and normalized expression matrix provides a “corrected” dataset which represents transcriptomic data of different single cells within the sample. It is however still fairly sparse and highly dimensional, containing thousands of genes per thousands of cells. To extract informative signals and reduce computational power, feature selection and dimensionality reduction are usually implemented. Feature selection processes identifies highly variable and informative genes within the dataset, and disregard genes that has universal or no expression across all cells such as housekeeping genes. Reduction of genes included in the expression matrix can drastically increase computational speed for downstream analysis, and is implemented by algorithms such as HVG based approach (Brennecke et al., 2013), FVG by Seurat (Satija et al., 2015) or dropout based approach M3Drop (Andrews & Hemberg, 2018). Dimensionality reduction is first achieved by Principal Component Analysis (PCA), where the data is linearly transformed into dimensions of principal components, such that the first components describe the most variance in the data. Higher dimensions in the PC space are then ignored, and performed with T-distributed stochastic neighbor embedding (t-SNE, van der Maaten & Hinton, 2008) or Uniform manifold approximation and projection (UMAP, Becht et al., 2018) to project data points onto 2 or 3 dimensional space for visualization.

Following these processes allow the data to be easily handled and visualized while preserving biological variations among different cells, thus allowing researchers to study the heterogeneity of their samples. Separating cells into meaningful subpopulation facilitates the identification of different cell types, thus different clustering methods are usually applied. Unsupervised clustering methods are typically used initially to identify cells that are similar to each other based on their expression patterns. These methods

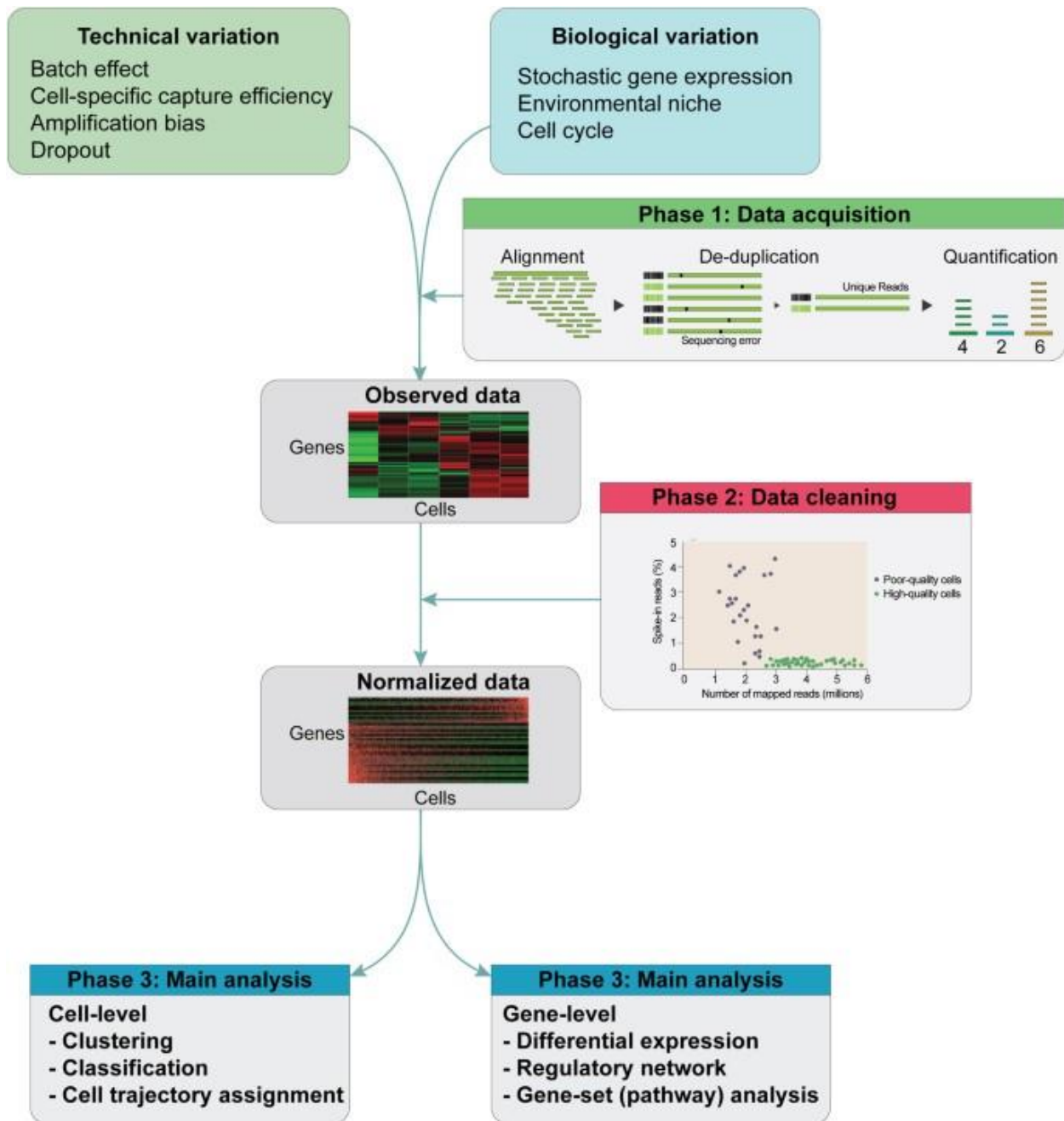


Fig1.3.2.3 A schematic overview of scRNA-seq analysis pipelines. scRNA-seq data are inherently noisy with bconfounding factors, such as technical and biological variables. After sequencing, alignment and de-duplication are performed to quantify an initial gene expression profile matrix. Next, normalization is performed with raw expression data using various statistical methods. Additional QC can be performed when using spike-ins by inspecting the mapping ratio to discard low-quality cells. Finally, the normalized matrix is then subjected to main analysis through clustering of cells to identify subtypes. Cell trajectories can be inferred based on these data and by detecting differentially expressed genes between clusters. (Hwang et al., 2018)

include k-means clustering, hierarchical clustering, density-based clustering and graph-based clustering (Andrews & Hemberg, 2018). scRNA-seq data specific methods such as SC3 (Kiselev et al., 2017) and Seurat (Satija et al., 2015) were also developed, allowing robust and accurate identification of cell clusters based on machine-learning algorithms. Once clusters are identified, differential expression of genes among different clusters can be determined, by comparing expression profile of individual clusters with the rest of the dataset. Significantly high expression of certain genes, known as markers, may signify their cell identity based on previous knowledge of cell markers. Thus, by identifying cell types using differential markers, different cell types can be obtained and further studied individually or comparatively. Novel markers could also be discovered this way to identify potential subpopulation of previously defined cell types.

1.4 Retinal Cell types

As our study focuses on the study of glaucoma, an eye disease which majorly affects the retina, understanding the role and function of the retina as a tissue, as well as the cell types that exist within the retina is essential for single cell study of the disease. The retina is the innermost layer of the eye where light reaches after passing through the lens, in which visual perception is formed and transmitted. This photosensitive tissue achieves its function by the complex harmony between multiple different cell types. Fig1.4 shows the schematic drawing of the major retinal cell types across the layers of the retina. The photosensitive cells, rods and cones, react to light stimulus at the outermost layer of the retina. Rod cells function mainly in low light intensity environment and provides black and white vision. The three types of cone cells function primarily in high light intensity environment and are particularly sensitive to low, medium, and high wavelength light which provides colour vision for the detection of red, green and blue light respectively, as well as allowing high accuracy vision. Visual signals are cascaded through the inner layers where amacrine, bipolar and horizontal cells perform primary processing before reaching the

retinal ganglion cells (RGCs). RGCs are neurons where the cell bodies are located near the inner surface of the retina, while their long axons bundle up to form the optic nerve and reaches the brain by passing through the optic nerve head. They are responsible for relaying visual information from the photoreceptors to the brain through action potentials, thus damages to the RGCs result in vision loss.

There are also three types of glial cells found in the human retina, Müller glia, astrocytes, and microglia (*Glial Cells of the Retina by Helga Kolb – Webvision*, n.d.). Müller glia is the most common type of glial cells found across the retina, providing both structural and functional support for neurons, including providing homeostasis, protection to excessive neurotransmitters and responds to injury (Yurco & Cameron, 2005). Astrocytes function similar to Müller glia and are found in vascularized region (Watanabe & Raff, 1988). Microglia represent the resident tissue macrophages and play important roles in retinal homeostasis and maintaining synaptic structure and function (Rashid et al., 2019; Silverman & Wong, 2018). However, trauma and insults to the retina trigger microglia's inflammatory response and function like macrophages, which could lead to enhanced phagocytosis and secretion of neurotoxins (Langmann, 2007).

Comprehensive studies of retinal cell types using single cell transcriptomic analysis was performed on both human (Menon et al., 2019) and mouse (Macosko et al., 2015a) retinæ. Other publications further utilized scRNA to identify subpopulation of bipolar cells (Shekhar et al., 2016), amacrine cells (Yan et al., 2020) and RGCs (N. M. Tran et al., 2019a), creating an extensive single cell transcriptomic atlas of the retina and provided different novel cell markers. By referencing their findings, future scRNA experiment of retinal tissues can easily identify cell type markers, facilitating the research of molecular changes and variations in different cell subpopulation during retinal development and disease progression. Further research opportunities may be present to studying different regions on the retina, as the proportions and distribution of cell type subpopulation may vary depending on the region of sample selected, such as closer to the optic nerve head, or further away near the peripheral regions. Studies related to glaucoma could focus on the optic nerve head, where astrocytes and RGCs are of the highest concentration.

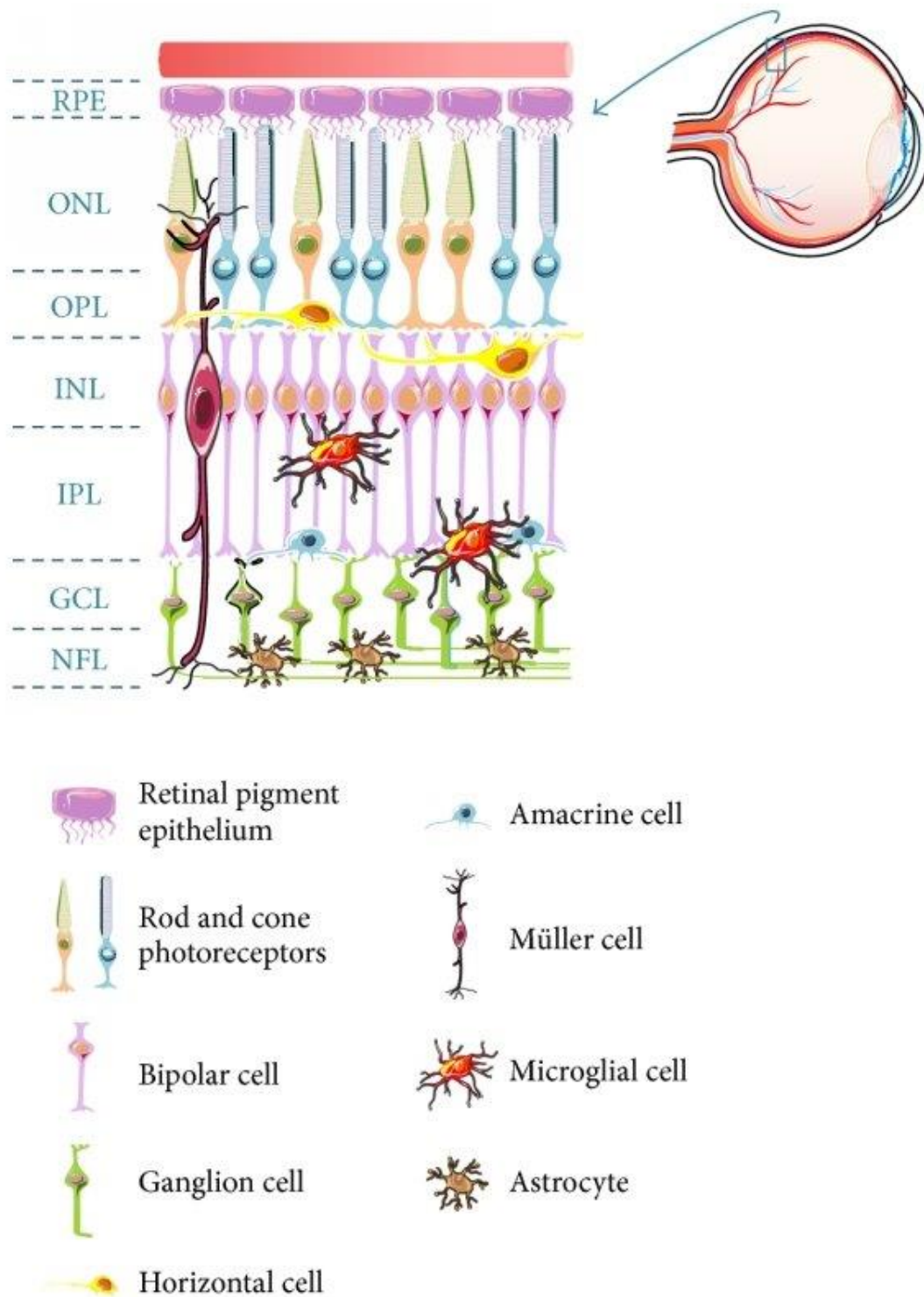


Fig1.4 Schematic representation of the major retinal cell types and their organization in the retina. (Madeira et al., 2015)

Chapter 2: Materials and Methods

2.1 Preface

This project aims at using single cell RNA-sequencing analysis to understand the pathogenesis of glaucoma, using gene knockout animal models. This chapter outlines the experimental methods used throughout the experimental pipeline of single cell experiment, including animal handling and tissue extractions to bioinformatics data analysis methods.

2.2 Mouse Methods

2.2.1 Animal housing and treatment

Animals involved in this study were obtained, housed, cared for, and used following the guidelines published by the Physiological Society of Japan, and with approval of the Animal Care Committee of Yamanashi University and University College of London. DBA/1J mice were used as WT samples, sourced from Japan SLC, INC. *ABCA1* knockout mouse model and Astrocyte-specific *ABCA1* knockout model were created using protocols suggested in Hamon et al., 2000 and Karasinska et al., 2013 respectively. Mice were maintained in a pathogen-free, temperature (23 °C)- and humidity (55%)-controlled facility with a 12-h (6:00 am–6:00 pm) light-dark cycle with free access to food and water.

2.2.2 Eye and Retina dissection

9-11 months old mice samples were euthanized using CO₂ chambers. Their eyeballs were dissected and placed in a glass slide with shallow volume of PBS. Using angled-forceps to hold the eyeball in position, a small spring scissor was used to create an incision at the cornea-sclera divide. The scissor then cuts along the divide around the eyeball until the cornea and iris is lifted and the lens is released. The retina is then gently pulled out of the eyecup using angled-forceps, carefully separated from the choroid layer, while gently separating at the fovea and preserving neurons at the optic nerve head.

2.2.3 Retina Dissociation

2.2.3.1 ABCA1 samples

The Worthington Papain Dissociation System was used to dissociate *ABCA1*^{-/-} experiment tissues into single-cell suspension. Briefly, the retinal tissue was placed in the mixture of 100 µL papain solution (20 units/ml with 1mM L-cysteine and 0.5 mM EDTA) and 5 µL of DNase solution (2000 units/ml). The mixture was incubated at 37°C for 15 minutes after gentle titration. The dissociated cell solution was then diluted with 900 µL of PBS and passed through a 35 µm cell filter. Dissociated samples were then counted using LUNA-FL™ Dual Fluorescence Cell Counter (Logos Biosystems) to verify quality of dissociation and viability of cell samples.

2.2.3.2 P2Y6 samples

Cold dissociation using Bacillus Licheniforms Protease (Sigma-Aldrich) were chosen to dissociate *P2Y6*^{-/-} experiment samples by our Japanese collaborators. Briefly, tissues were digested in 1mL of protease solution (5 mM CaCl₂, 5 mg/ml Protease, 62.5 U/ml DNase, in 2.0 ml tube) for 10 min on ice with shaking, Samples were then pipetted and shaken on ice, then filtered with a 70 µm filter and washed with PBS twice with 1mL each. Samples were then centrifuged at x300 g for 7 min at 4°C, removed with supernatant and refill with PBS twice, which results in a dissociated sample that is then inspected using the Trypan blue assay.

2.2.4 10X Single Cell Chromium system

Dissociated samples were loaded onto the 10X Chromium controller (10X Genomics), and single cell RNA-seq libraries were prepared using the Chromium Single Cell 3' reagent kits v3. In accordance with the manufacturer's protocol, single cells are captured in nanoliter-scale Gel Beads-in-emulsion (GEMs) where cells were lysed. cDNA libraries of each cell were constructed by reverse transcription of mRNAs from that were encapsulated within a GEM, and tagged on the 3' end with primers including a 16bp cell-specific 10X Barcode and a 12bp read-specific unique

molecular identifier (UMI). GEMs were then broken up and cDNAs from all cells in a sample were pooled, purified and amplified via Polymerase Chain Reaction (PCR). Library quality was ensured by using a high-sensitivity D5000 screentape and Qubit 2.0 fluorometer. Each sample is then sequenced on the NextSeq 500 systems (Illumina) for paired-end sequencing.

2.2.5 Genotyping

Mouse genotyping was performed in accordance with the protocols provided by our Japanese collaborators, which utilizes the KAPA2G Fast PCR Kit. Mouse tail of 2-3mm was placed into a safe lock tube of 1.5mL, added with 200µl of 50 mM NaOH/H₂O and incubated for 15-30mins at 100°C dry bath. 22 µl of 1M Tris-HCl (pH8.0) was added and mixed with the solution, and centrifuged at 12,000 rpm for 1 minute at 25°C. Stock solutions of 100µM primers in TE buffer were synthesized by Thermo scientific, and working solution were prepared by diluting with TE buffers to make 10µM working solutions. Table2.2.5.1 shows the sequence of the primers for WT and KO amplicons, Table2.2.5.2 shows the PCR reaction mixture formula, and Table2.2.5.3 details the PCR reaction procedures. Resultant sequences should produce 750bp and 540bp in length for WT and KO samples respectively, by running the PCR result samples on a 2% agarose gel with ethidium bromide at 100 and visualized using a UV transilluminator.

| <i>Primer</i> | <i>Sequence</i> |
|--------------------|-----------------------|
| <i>WT forward</i> | TGGGAACCTCCTGCTA AAAT |
| <i>WT reverse</i> | CCATGTGGTGTGTAGACA |
| <i>KO forward</i> | TTTCTCATAGGGTTGGTCA |
| <i>KIO reverse</i> | TGCAATCCATCTTGTTCAT |

Table2.2.5.1 PCR primer sequences.

| <i>Reagent</i> | <i>Volume (μl)</i> |
|---------------------------------------|--------------------|
| <i>KAPA2G Fast Hot Start Readymix</i> | 7.5 |
| <i>10μM primer solution</i> | 0.3 |
| <i>genomeDNA template</i> | 1 |
| <i>H₂O</i> | 5.3 |
| <i>Total</i> | 15 |

Table2.2.5.2 PCR reaction mixture.

| <i>Steps</i> | <i>Temperature(°C)</i> | <i>Time(s)</i> | <i>No. Cycles</i> |
|--------------|------------------------|----------------|-------------------|
| 1 | 95 | 180 | 1 |
| 2 | 95 | 15 | 30 |
| 3 | 55 | 15 | |
| 4 | 72 | 3 | |
| 5 | 72 | 60 | 1 |

Table2.2.5.3 PCR thermocycle.

2.3 Data analysis methods

2.3.1 Cell ranger

Cell Ranger is a set of programs developed by 10X Genomics to process sequenced data from Chromium single- cell experiments. Upon receiving raw base-calling files (BCL) from the output of Illumina sequencing machines, *CellRanger mkfastq* was applied to demultiplex sequencing reads into FASTQ files. *CellRanger count* then performs splicing-aware alignment (Dobin et al., 2013) of reads to the GRCm38 mouse genome and UMI counting to remove PCR duplicates. The program then employs the EmptyDrops method (L Lun et al., n.d.) to perform cell-calling based on levels of RNA content within each cell. The output of *CellRanger count* includes a web summary HTML file which includes basic statistics about the sample run, as well as both “cell-only” and “unfiltered” sample feature count matrices. The cell-only filtered feature count matrix in

the form of 3 tabular files (*barcodes* – representing columns of individual single cell labelled using unique 10X barcodes, *features* – representing rows of gene identifiers, and *matrix* – containing the sparse matrix of the feature counts) were carried over to R Studio for further analysis.

2.3.2 Seurat

The R toolkit package Seurat (v3.1.4; <https://github.com/satijalab/seurat>) was utilized for downstream analysis.

2.3.2.1 Individual samples of ABCA1

Firstly, individual samples of ABCA1 experiment were inspected. Seurat object were created in *RStudio* by loading in feature count matrices from the *CellRanger* pipeline with a pre-filtering of minimum 200 features. Gene expression is then “Log-Normalized” by log-transforming the product of gene count by a scale factor of 10,000, and linearly scaled per gene such that the expression mean across all cells is 0 and that the variance across all cells is 1. The top 2000 “highest-variable” genes were then chosen for subsequent analysis, by ranking the standardized variance obtained from a variance-stabilizing transformation described in Stuart et al., 2019. Principal Component Analysis (PCA) was then applied as a tool for linear dimensional reduction, as it computes principal components in lower dimensions while preserving data variation. The top 15 principal component were then used for Seurat’s graph-based clustering approach (Macosko et al., 2015b). In brief, the clustering steps involve the construction of a K-nearest neighbour (KNN) graph with which the edges are refined using Jaccard similarity and optimized using the Louvain algorithm. t-Distributed Stochastic Neighbour Embedding (t-SNE) was applied to visualize the clustered dataset by performing non-linear dimensional reduction (van der Maaten & Hinton, 2008). Due to the abundance of rod cells present in the retina, clusters expressing high levels of rod-markers such as *Rho*, *Sag*, *Gnat1* or *Pde6a* were filtered to create “Non-rod” subclusters, which were then separately clustered and projected.

2.3.2.2 ABCA1 sample integration

Gene expression integration was then performed to allow meaningful comparison between sequencing data of different biological samples. 3 WT and 5 ABCA1 KO samples were integrated using methods implemented in Seurat's toolkit, as described in Stuart et.al. Firstly, individual samples were log-normalized and feature extracted with the same parameters aforementioned (Chapter 2.3.2.1). *FindIntegrationAnchors* function was applied to a list of all sample data, to identify pairs of cells originating from different datasets that suggest to having matching biological state. These cells act as “anchors” between datasets, in which *IntegrateData* function then corrects expression levels by removing technical differences and integrate all cell data into a single Seurat object (Stuart et al., 2019b). Scaling, PCA and Clustering were then performed on the integrated object as mentioned before, but using the top 30 principal components to account for the large amount of cells and expression level variation. Uniform Manifold Approximation and Projection (UMAP) was used as a non-linear dimensional reduction technique for visualization, as it performs better than tSNE when applied to datasets with much more cells. *FindAllMarkers* function was then applied to perform Wilcoxon Rank Sum test on each cluster, identifying differentially expressed genes within each cluster in comparison to the rest of the dataset. Retinal cell-type specific markers were then identified in the differential expression gene lists, allowing the labelling of each clusters.

2.3.2.3 ABCA1 sub-clustering

With the interest of studying Müller Glia/ astrocytes, immune cells and RGCs in the context of this study, clusters expressing markers of these cell types were isolated and inspected separately. Clusters with high expression of *Apoe*, *Clu*, *Glul* and *Gpr37* formed the Müller Glia/ astrocyte subclusters, while cluster with high expression of *C1qa* and *P2ry12* were suspected to contain immune cells such as microglia. Cluster with expression of *Chrn3*, *Slc17a6*, *Thy1* and *Chrna6* were labelled as “mixed group” as it contains a mixture of different cell types while containing cells

expressing RGC markers. PCA, clustering and tSNE were applied to segregated cell clusters respectively, and differential expression of subclusters were identified using *FindAllMarkers* function.

2.3.2.4 P2Y6 samples pre-processing: SoupX

3 WT and 3 P2Y6 KO mouse retinal tissue were processed by our collaborators in the University of Yamanashi, using similar 10X Genomics Single Cell Chromium pipeline as the ABCA1 experiment. *CellRanger* outputs of “cell-only” and “unfiltered” sample feature count matrices were provided for further analysis.

Pre-filtering of background contamination was performed using R package *SoupX* (v1.5.2) using a novel method developed and implemented by Young & Behjati, 2018. Briefly, the *SoupX* algorithm estimates background RNA expression profile from non-cell droplets, obtained from exclusive cells in “unfiltered” feature matrix that were not in the “cell-only” matrix. *SoupX* then estimate (or manually determine) the fraction of contamination present in the “cell-only” feature matrix, attempting to distinguish between RNA expression that originates from actual cell content and RNA expression that belongs to background contamination. *SoupX* then corrects the expression profile using the background contamination expression profile and contamination fraction, producing a modified feature matrix that can be analysed in subsequent pipeline. Individual samples of the P2Y6 experiment were thus cleaned-up using *SoupX*, and proceeded with the aforementioned pipeline of Log-normalization, feature extraction, scaling, PCA, clustering and UMAP using *Seurat*.

2.3.2.5 P2Y6 samples integration

SoupX corrected feature matrices of 3 WT and 3 P2Y6 KO retinal samples were integrated using the same pipeline as mentioned in Chapter 2.3.2.2. Due to the large amount of cells included in the samples, reciprocal PCA (RPCA) (Hao et al., 2021) was used for the *FindIntegrationAnchors*

function instead of the default canonical correlation analysis (CCA) (Stuart et al., 2019b) method. The top 25 principal components were chosen for linear dimensional reduction.

2.3.2.6 P2Y6 samples subclustering

Similar to ABCA1 experiment, cell types of interest: RGCs and glial cells were selected for sub-clustering. Clusters with high expression of RGC markers (*Nefl*, *Thy1*, *Stmn2*, *Resp18*) and glial cell markers (*Glul*, *Dbi*, *Clu*, *ApoE*) were isolated as separate subset of the integrated data, and processed with PCA, clustering, and UMAP projection. Differential expression with each subclusters were identified using the *FindAllMarkers* function.

2.3.3 Ontological Analysis

Differential expression gene lists generated from *Seurat* were usually followed with ontological analysis using the Ingenuity Pathway Analysis software (Qiagen). “Canonical Pathways” and “Disease and Functions” list can be obtained, with p-values signifying the relevance and significance of the related pathways, while z-score relates the degree of activation (+) or deactivation (-) of said pathways or mechanisms.

Chapter 3: ABCA1^{-/-} Experiment

3.1 Preface

The pathogenesis of glaucoma is poorly understood despite it affecting millions of the population around the world. Although the phenotype of the disease is defined as the progressive degeneration of the retinal ganglion cells (RGCs) on the retina, many studies have suggested roles and involvement of other cell types. The use of experimental animal models that show glaucoma phenotype would benefit studies in understanding the mechanisms activated during disease development, and changes in survivability or behaviour of different cell types.

In this chapter, we utilized our Normal Tension Glaucoma (NTG) mouse model which knocked out the gene *ABCA1*, developed and provided by our collaborators from the University of Yamanashi in Japan, and performed single cell RNA-seq analysis to gain insight and understanding of different cell type behaviours and transcriptomic changes. The full experimental pipeline of single cell analysis was implemented, starting from sample preparation to bioinformatics data analysis. Computational analysis allowed for the identification of different cell types, extracting differential expressions between wild type (WT) and knockout (KO) samples, and ontological analysis to predict activated canonical pathways.

3.2 Tissue sample preparation

3.2.1 Genotyping

Wild type (WT) and *Abca1*^{-/-} (KO) mice were transported from our Japanese collaborator to UCL animal facilities. Prior to preparing samples for single cell analysis, genotype of different mice was verified using the protocols mentioned in Chapter 2.2.5. Fig3.1 displays the resulting gel imaging, with 750bp fragments detected for WT samples and 540bp detected for KO samples.

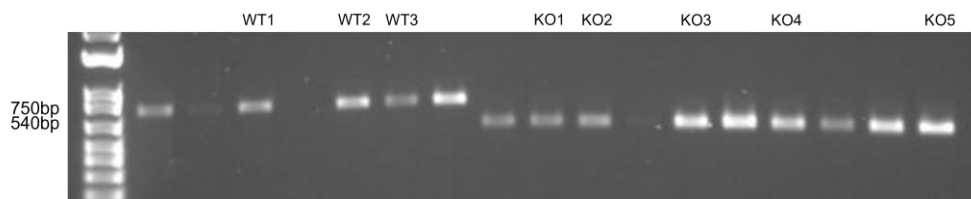


Fig3.1: PCR genotyping results for 3 WT and 5 KO samples.

3.2.2 Tissue dissociation

3 WT and 5 KO mouse were euthanized, and their retinal tissue were dissected at the UCL Animal facilities using protocols as described in Chapter2.2.3.1. To maintain cell viability, intact retinal tissues were placed in separate tubes of PBS on ice and transferred to the laboratory at QMUL Genome Center, for sequencing by the 10X Genomics Chromium Machine. The tissues were then dissociated on-site using the Worthington Papain Dissociation System, filtered and treated with the pre-loading procedures recommended by 10X Genomics. The following table summarizes the final loading conditions of each sample, and the images display the cell conditions under the LUNA-FL™ Dual Fluorescence Cell Counter. Green fluorescent staining patterns signify fresh living cells. All dissociated samples achieved satisfactory viability (>90%) and live cell concentration. The targeted number of cells was 4,000 per sample, and triple the number of cells were loaded onto each lane in accordance to the recommendations by 10X Genomics technicians, supported by their experience with machine uptake from other previous samples.

| Sample | Total cell concentration (1e6 cells/mL) | Live cell concentration (1e6 cells/mL) | Dead cell concentration (1e5 cells/mL) | Viability (%) | Average cell size (µm) |
|---------------|--------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|----------------------|-----------------------------------|
| WT1 | 2.43 | 2.32 | 1.10 | 95.5 | 5.6 |
| WT2 | 4.01 | 3.81 | 2.02 | 95 | 7.3 |
| WT3 | 4.17 | 3.93 | 2.42 | 94.2 | 7.9 |
| KO1 | 3.43 | 3.39 | 4.46 | 98.7 | 8.1 |
| KO2 | 1.03 | 0.993 | 0.442 | 95.9 | 7.6 |
| KO3 | 2.46 | 2.33 | 1.31 | 94.7 | 7.9 |
| KO4 | 4.82 | 4.6 | 2.21 | 95.4 | 8.4 |
| KO5 | 4.09 | 3.89 | 1.95 | 95.2 | 6.8 |

Table3.2.2: Loading conditions of dissociated retinal cells

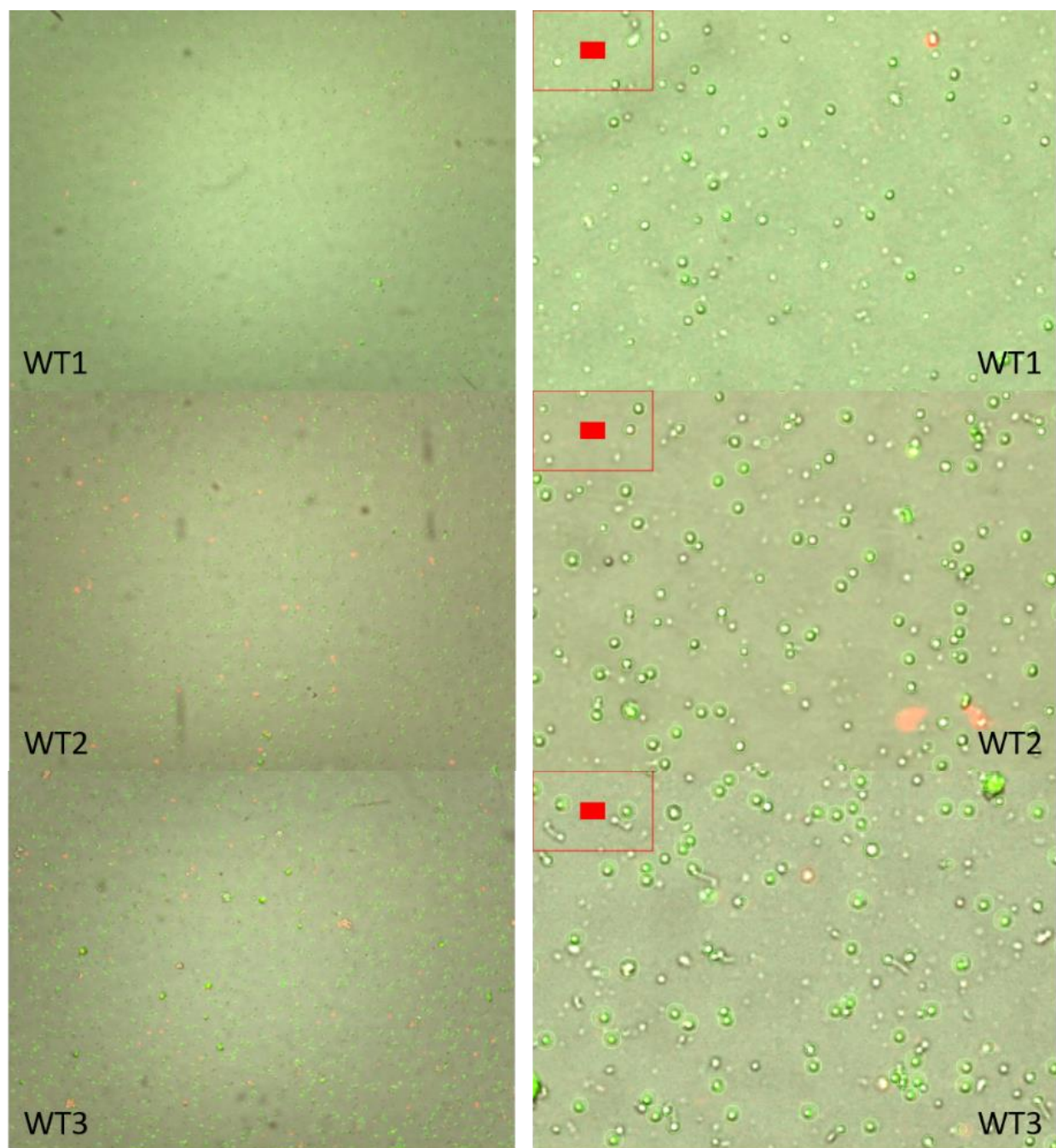


Fig3.2.2.1: Fluorescent Cell counter images of three dissociated wild type mouse retinal cells

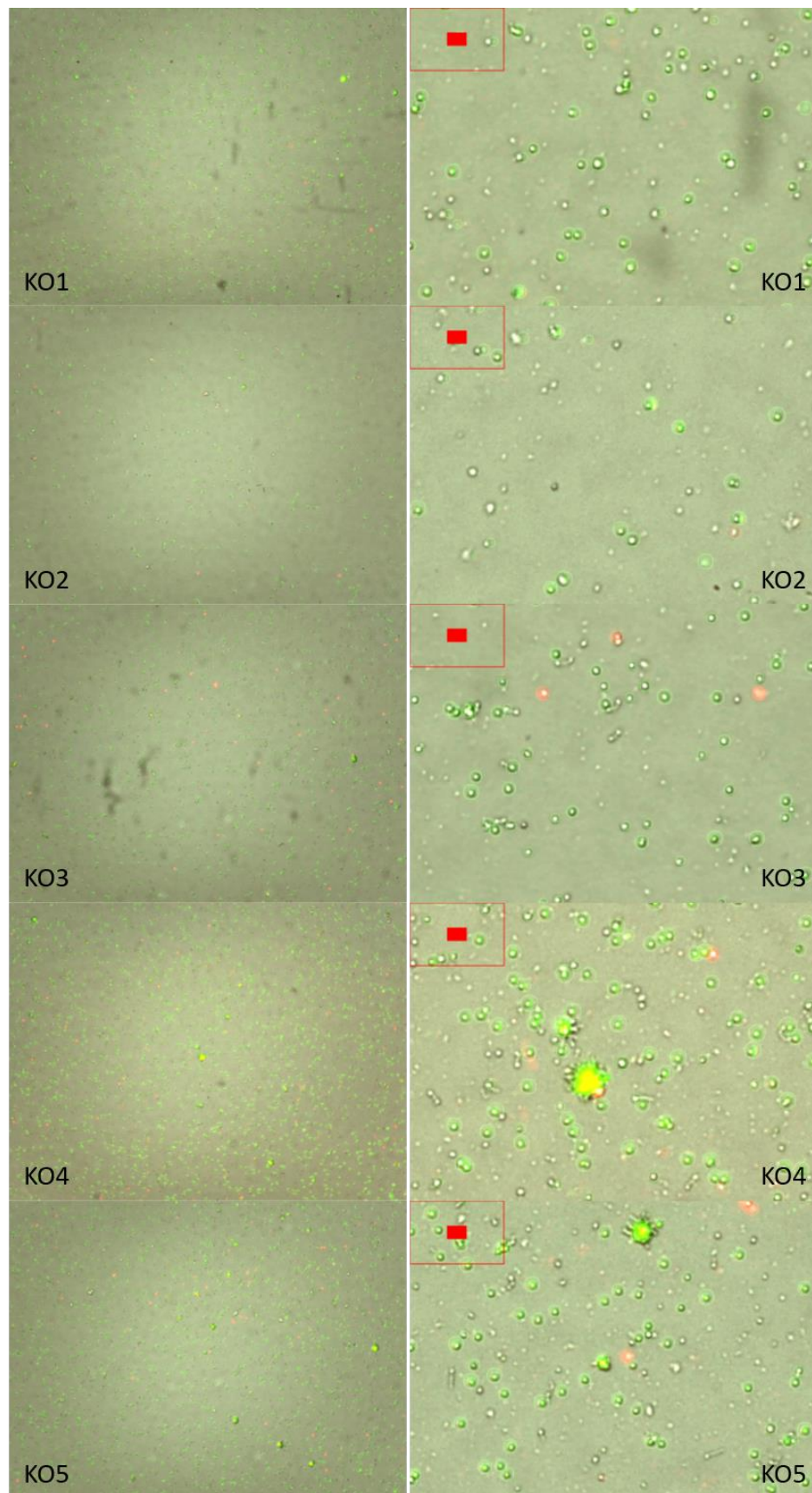


Fig3.2.2.2: Fluorescent Cell counter image of five dissociated ABCA1-/- mice retinal cells

3.3 Data Analysis

3.3.1 10X Genomics CellRanger output

The cDNA library of 8 retinal tissue samples constructed by the 10X Genomics Chromium Machine were sequenced into digital data. Sequencing data output was then processed through the CellRanger pipeline as described in Chapter 2.3.1, where sequencing reads were demultiplexed and aligned to the mouse genome. Feature-barcode matrices were then generated, which describes the expression profile of sequencing reads captured within each GEL which shares the same cellular barcode. CellRanger then employs the cell-calling algorithm based on the EmptyDrops method, to determine whether the expression profile described by reads sharing the same barcode is likely to originate from a single cell, rather than from an empty GEL or a doublet (GEL that encapsulated two or more cells).

By calling the default algorithm, the initial estimated number of cells average around 2,500 per sample (Fig3.3.1A). This is significantly lower than the target that was intended, and the variation among different samples in estimated number of cells and mean reads per cells were substantial. This is likely due to photoreceptors being low-expressing cells, as their sole function was to detect and transmit optical signals. Due to photoreceptors being overly abundant in the retina, majority of cells from the retinal tissues will be regarded as low-expressing cells when compared to other cell types in other tissues. Since the default algorithm attempts to remove low-expressing barcodes as they usually represent empty GELs with background RNA contamination, it is suggested to have underestimated the number of cells in retinal tissue samples. Supported by their past experience working with low-expressing datasets, the technicians from 10X Genomics forced the top 9,000 most expressing barcodes to be defined as cells in all samples. By slightly overestimating the number of cells, this allowed more room for filtering in subsequent analysis. Fig3.3.1B-C shows the mean reads per cell and median number of genes per cell from both the default cell-calling algorithm and force-9000 setup, with the latter having a more consistent and comparable mean reads per cell across all eight samples.

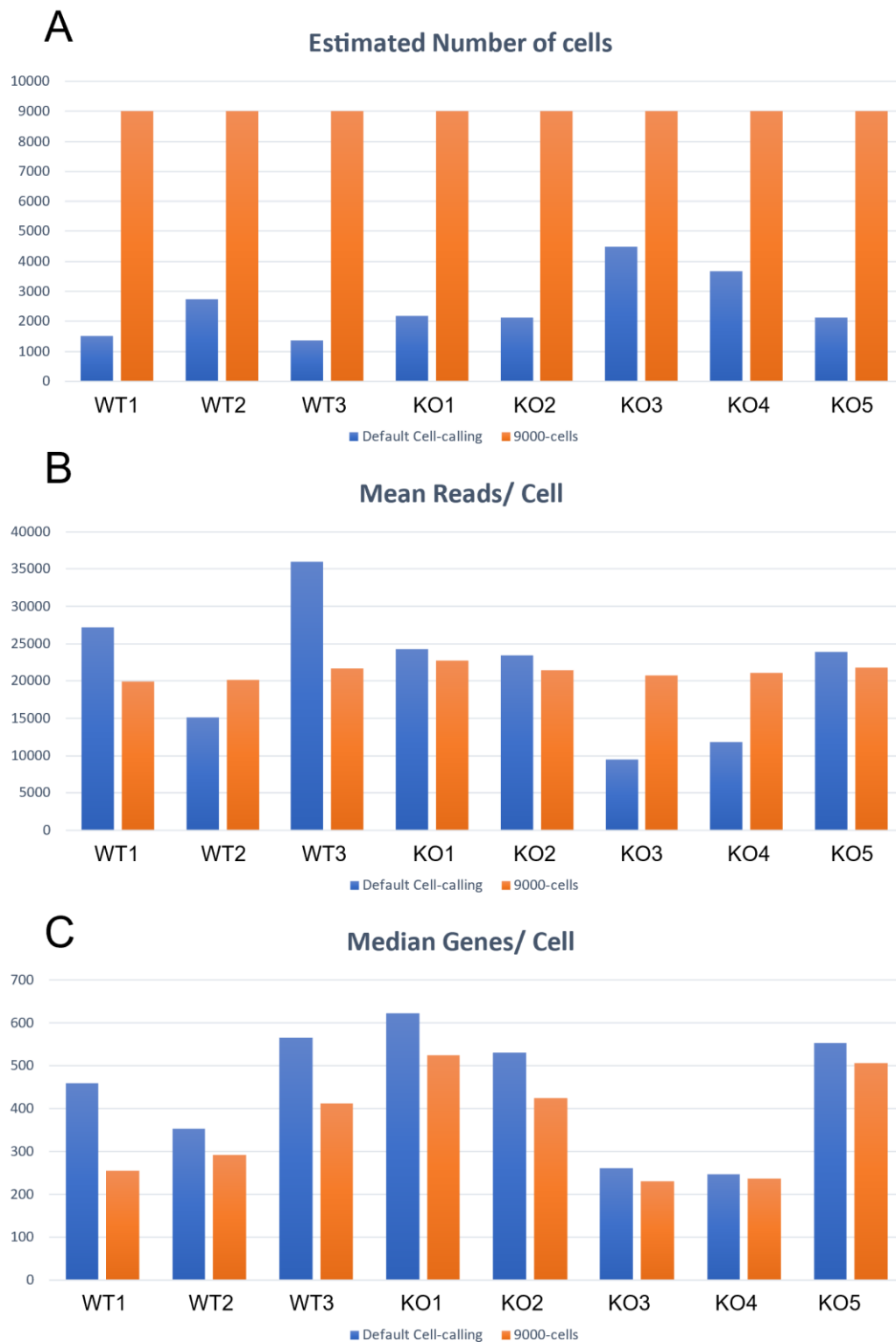


Fig3.3.1A) Estimated number of cells, B mean reads per cell and C) median number of genes per cell in ABCA1^{-/-} samples. Blue bars show statistics of using the default cell-calling algorithm of CellRanger, while orange bars show that of forcing each sample to include 9000 cells.

The following table summarizes the output from CellRanger for our 8 samples:

| Sample | Default cell-calling number of cells | Forced-9000 number of cells | Mean Reads/Cell | Median Genes/Cell | Sequencing Saturation (%) | Reads Mapped to Genome (%) | Total genes detected |
|------------|--------------------------------------|-----------------------------|-----------------|-------------------|---------------------------|----------------------------|----------------------|
| WT1 | 4483 | 9000 | 20719 | 231 | 88.1 | 95.7 | 16109 |
| WT2 | 3676 | 9000 | 21126 | 236 | 88.5 | 94.5 | 16587 |
| WT3 | 2113 | 9000 | 21756 | 506 | 77 | 93.7 | 17860 |
| KO1 | 1512 | 9000 | 19914 | 256 | 88.1 | 94.8 | 17272 |
| KO2 | 2735 | 9000 | 20116 | 291 | 86 | 94.2 | 17138 |
| KO3 | 1369 | 9000 | 21677 | 413 | 82.8 | 93.6 | 17345 |
| KO4 | 2169 | 9000 | 22722 | 524 | 79.5 | 93.2 | 17905 |
| KO5 | 2115 | 9000 | 21419 | 425 | 81.1 | 93.7 | 17308 |

Table3.3.1 CellRanger sequencing output summary of ABCA1-/- experiment samples

The sequencing results were highly satisfactory, as all the samples achieved over 90% of reads alignment to the mouse genome. Cells on average have around 20,000 reads in all samples, while the median genes per cells are generally less than 600 even though the sequencing saturation is about 80% or above. Contrasting to GELs that capture no cells and only the background RNA contamination, which would have low number of reads, genes and sequencing saturation, this data supports the premise that majority of the cells in the retina, mainly photoreceptors, has a low variety of genes in their expression profile. However, the total genes detected in all samples exceeds 16000, suggesting other cell types in the retina which have a more diverse expression profile were detected.

3.3.2 Seurat Individual samples

The eight individual samples were first inspected separately to try discovering different cell-types population and its distribution. Individual sample outputs from *CellRanger* creates the expression profile matrix of all cells, which were loaded into *RStudio* as *Seurat* objects using the *Seurat* package. They are then processed through the individual sample pipeline as suggested in Chapter 3.3.2.1, involving normalization, extraction of top 2000 features and scaling. PCA is then applied for dimensional reduction, and the top 15 Principal Components (PCs) were chosen for subsequent analysis, supported by the suggestion from an “Elbow plot”. The Elbow plot aids in deciding how many PCs were needed to capture majority of the data’s variation, by plotting the standard deviation within each PC against increasing number of PCs. As suggested from the example Elbow plot of WT1 in Fig 3.3.2.1, increasing PCs above 15 marginally increases the variance captured in the data, thus 15 PCs were sufficient.

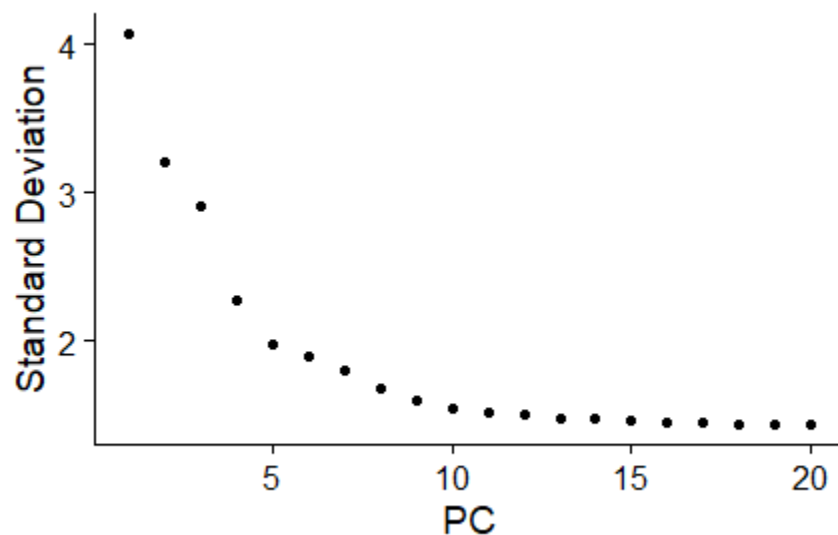


Fig3.3.2.1 Example Elbow plot of WT1 sample

Clustering and tSNE were then applied to visualize the 8 samples data in two-dimensional clusters, as shown in Fig3.3.2.2-9. Projections from all samples present similar phenomenon, where a large number of cells occupy the biggest cluster or area in the plot, while small individual clusters branches off.

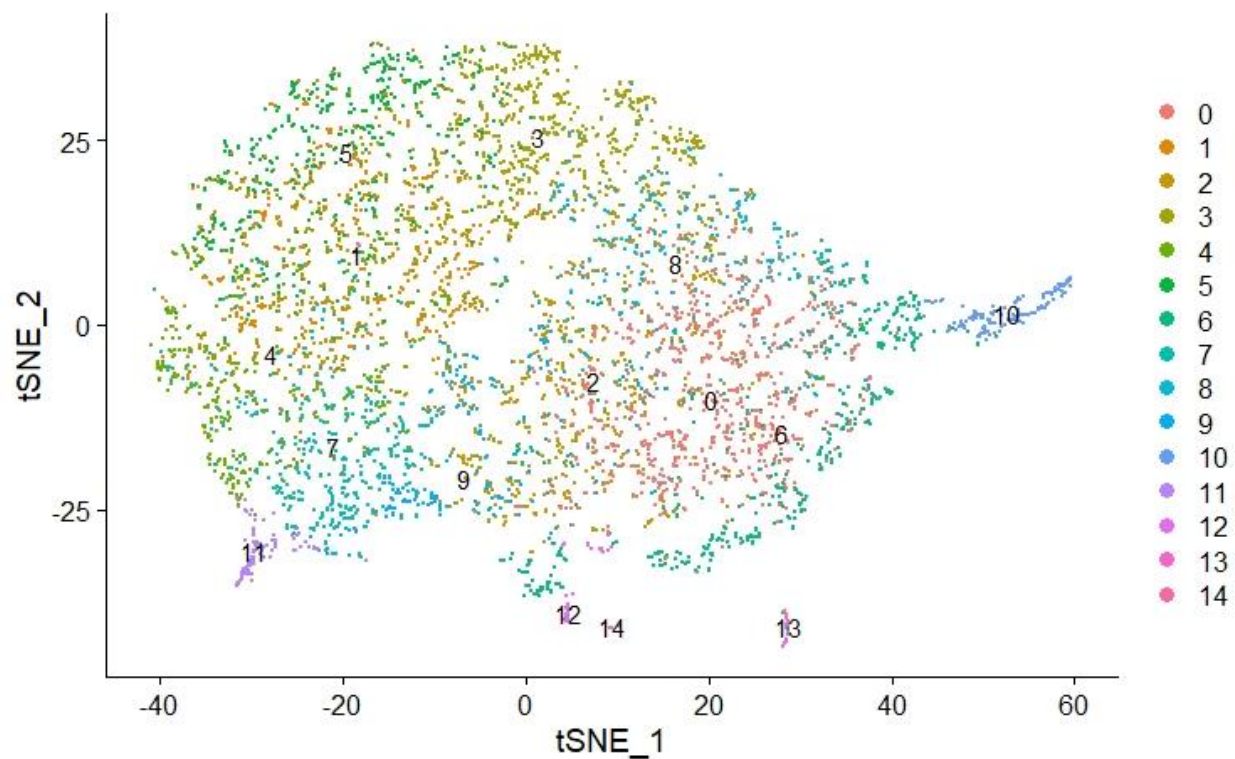


Fig3.3.2.2: WT1 tSNE projection

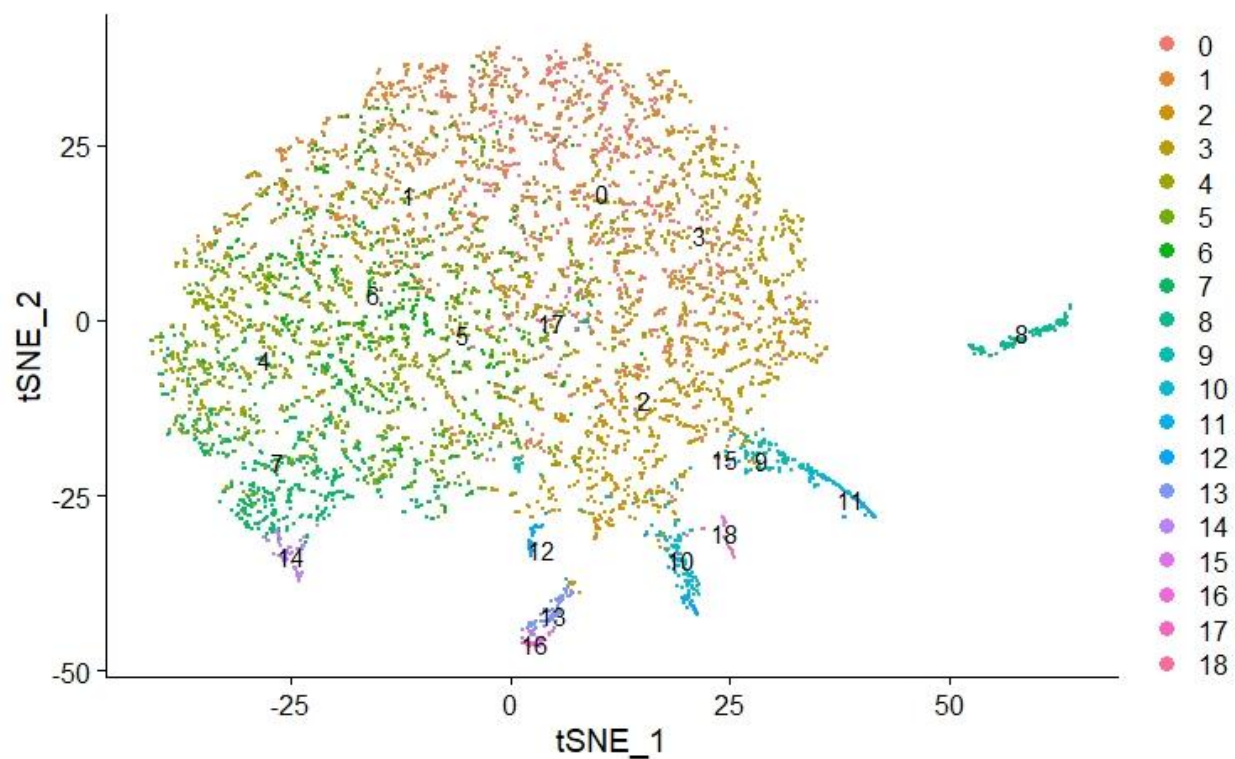


Fig3.3.2.3: WT2 tSNE projection

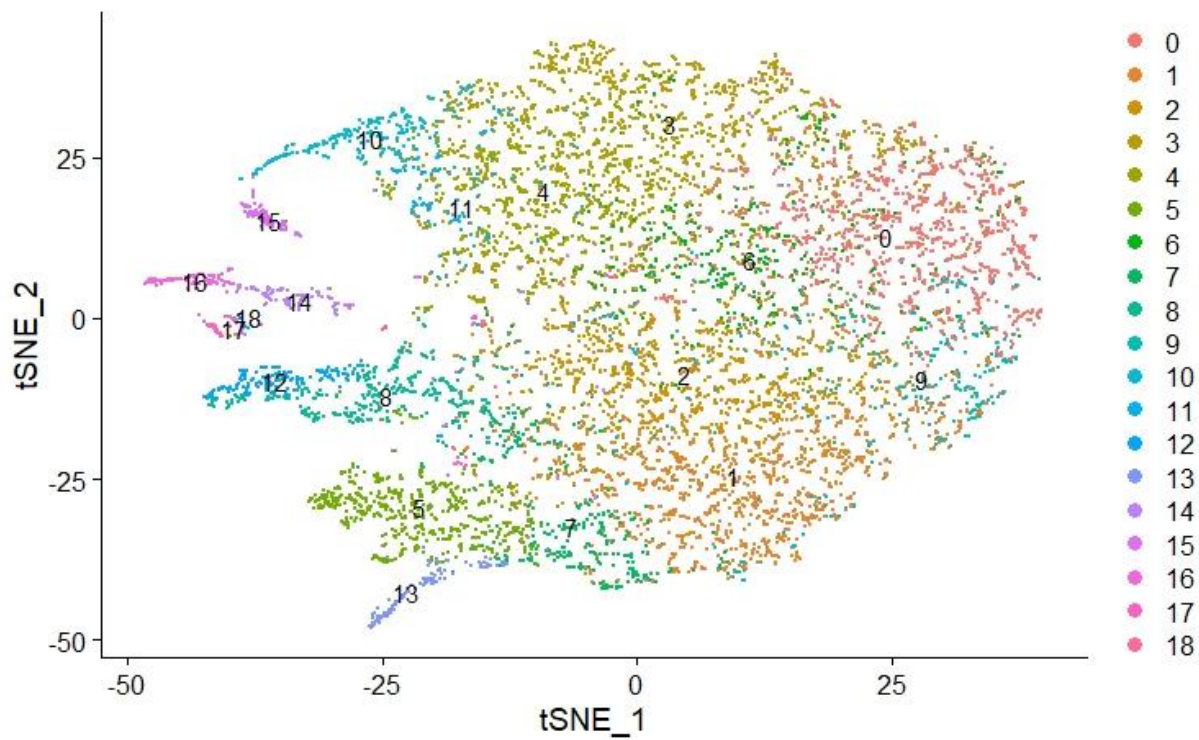


Fig3.3.2.4: WT3 tSNE projection

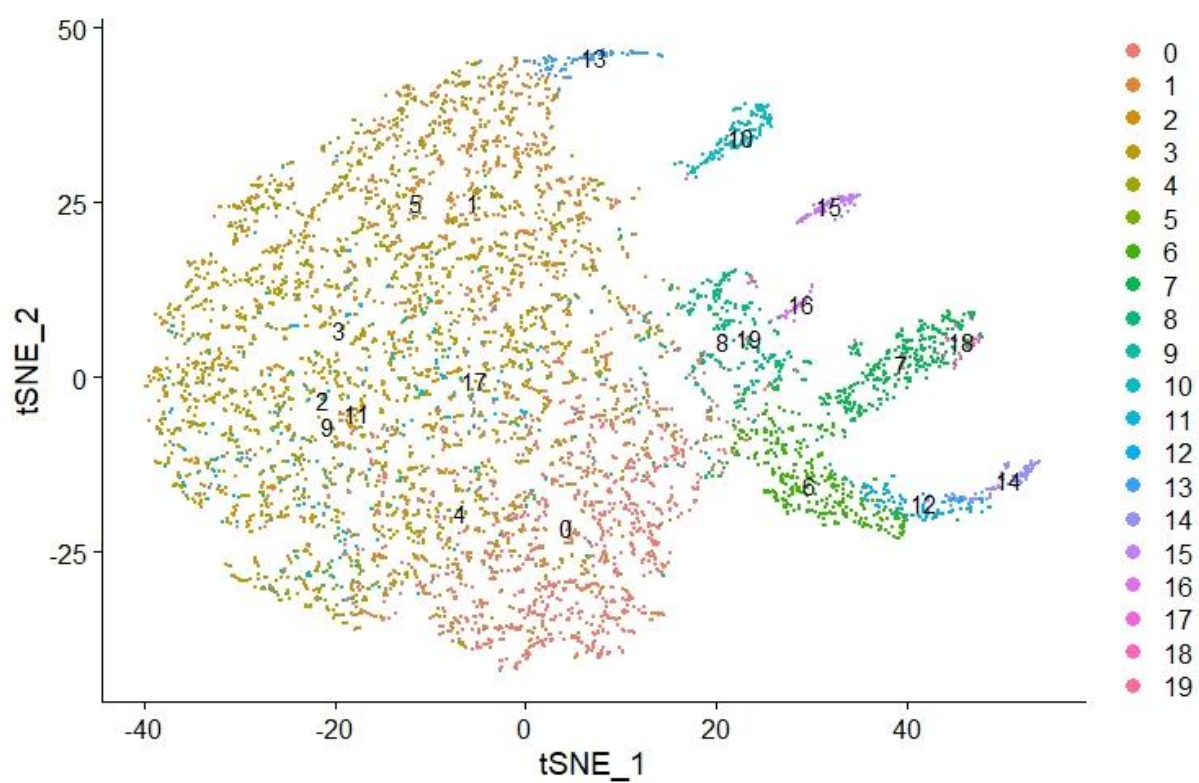


Fig3.3.2.5: KO1 tSNE projection

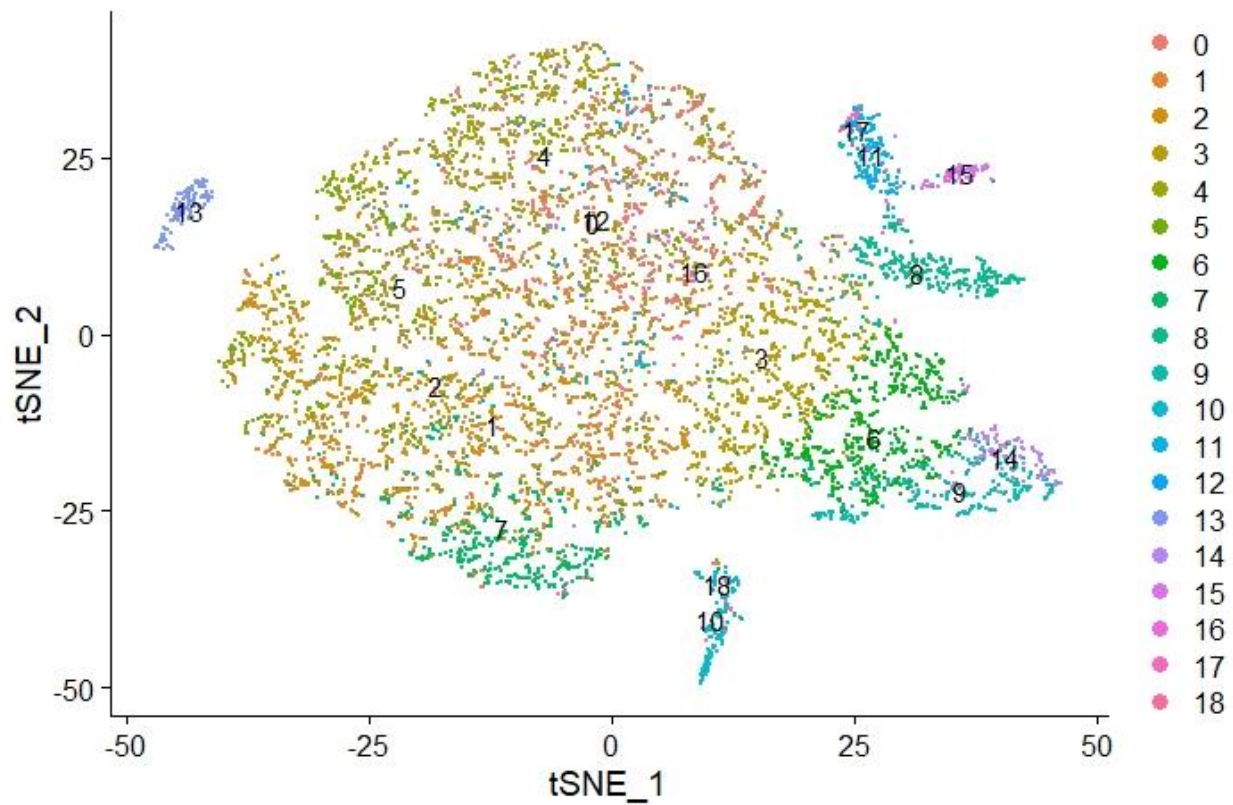


Fig3.3.2.6: KO2 tSNE projection

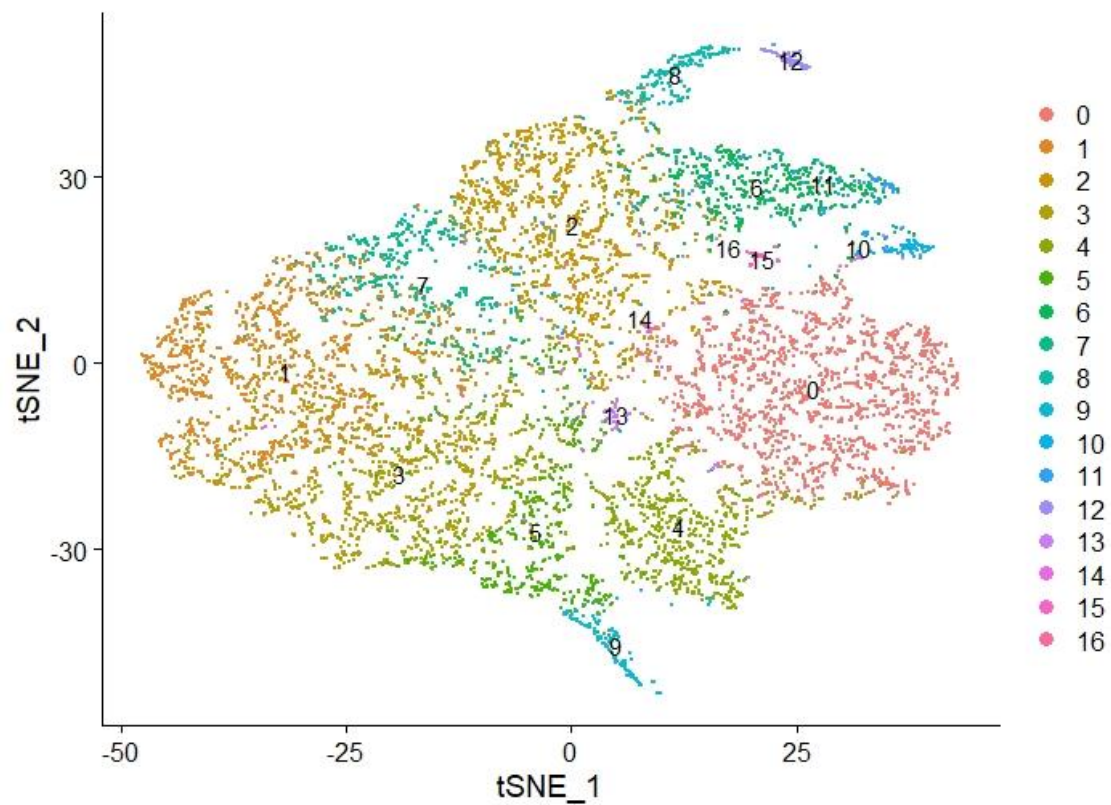


Fig3.3.2.7: KO3 tSNE projection

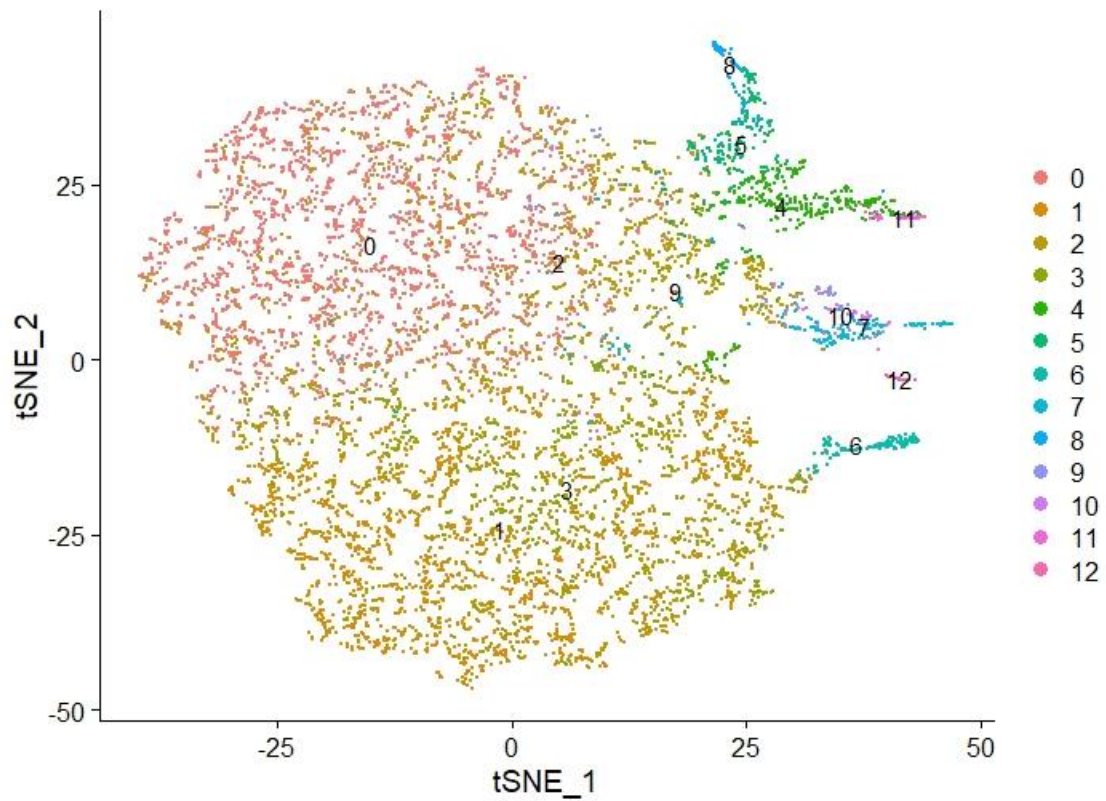


Fig3.3.2.8: KO4 tSNE projection

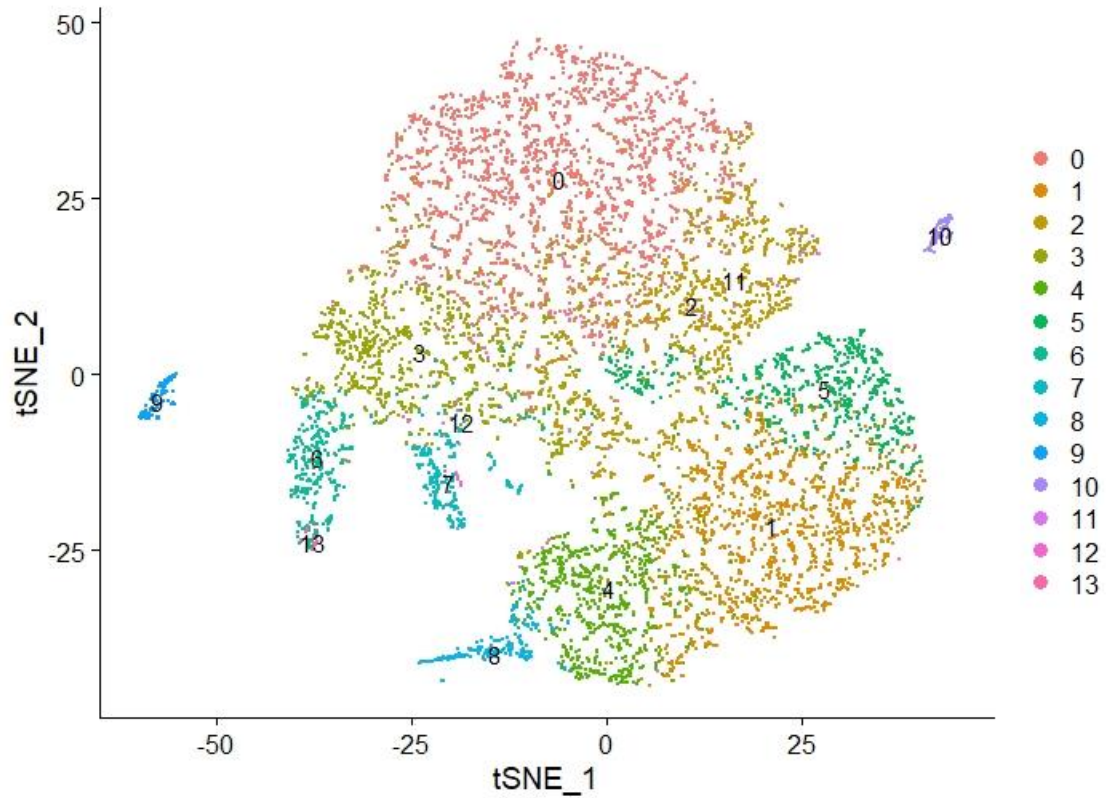


Fig3.3.2.9: KO5 tSNE projection

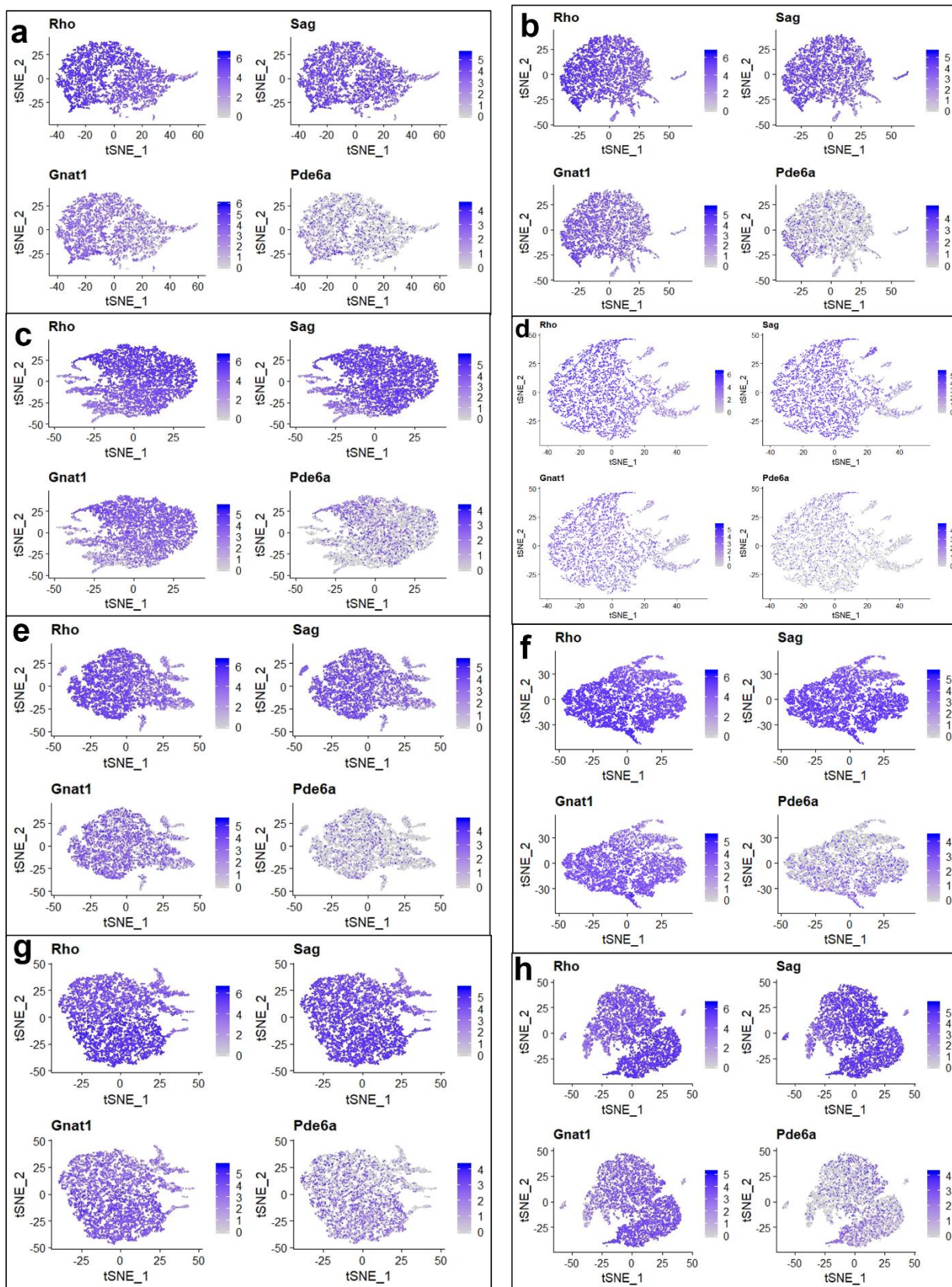


Fig3.3.2.10: Rod cell markers expression in sample (a) WT1, (b) WT2, (c) WT3, (d) KO1, (e) KO2, (f) KO3, (g) KO4, (h) KO5

Upon inspection of known markers such as *Rho*, *Sag*, *Gnat1* and *Pde6a*, the big clusters can be classified as rod photoreceptor, which concur with the known fact that this cell type occupies the majority of the cell population in the retina. To facilitate the identification of other cell types in a higher definition, cell clusters with relatively low expression of rod photoreceptor markers were separated as the “non-rod” population. PCA, clustering and tSNE were applied onto the “non-rod” subclusters, creating projection plots as shown in Fig3.3.2.11-18. Differentially expressed genes for each subcluster were discovered using the *FindAllMarkers* function, allowing us to label subclusters by identifying high expression of different cell type markers. The following table demonstrates the genes used for labeling each major retinal cell types, sourced from various supported literatures.

| Cell Type | Gene Markers |
|------------------------------------------|------------------------------------------------|
| Amacrine cells | Tfap2b, Gad1, Lrrn3, Gad2 |
| Astrocyte | Ndr2, Aqp4, Slc6a11, Aldoc, Apoe |
| Bipolar cells | Gng13, Vsx2, Prkca, Isl1, Sebox, Scgn |
| Bipolar - Cone | Tac3, Vsx1, Gja7 |
| Bipolar - Rod | Glr3, Grm6, Pcp2, Prkca |
| Cone cells | Arr3, Gngt2, Opn1sw, Cngb3, Gnat2 |
| Endothelial cells | Cldn5, Itgb1, Cxcl12, Pecam1 |
| Horizontal cells | Calb1, Onecut1, Onecut2 |
| Lens | Cryga, Crygf, Lenep |
| Immune cells (eg. Microglia/ Macrophage) | P2ry12, C1qa |
| Müller Glia | Clu, Dkk3, Glul, Vim, Abca8a |
| Microglia | Hla-dpa1, Hla-dpb1, Hladra |
| Photoreceptor Progenitor | Arrb1, Prdm1 |
| Progenitors | Hes1, Hes5, Pax6, Vsx2 |
| Retinal Ganglion Cells | Nefl, Stmn2, Vsnl1, Thy1, Resp18, Pou4f1, Opn4 |
| Rod cells | Rho, Sag, Gnat1, Pde6a, Cnga1, Nrl |
| Retinal Pigment Epithelium | Rgr, Rpe65, Rrh, Ttr |

Table3.3.2.1: Retinal Cell Markers used to identify and label subtypes

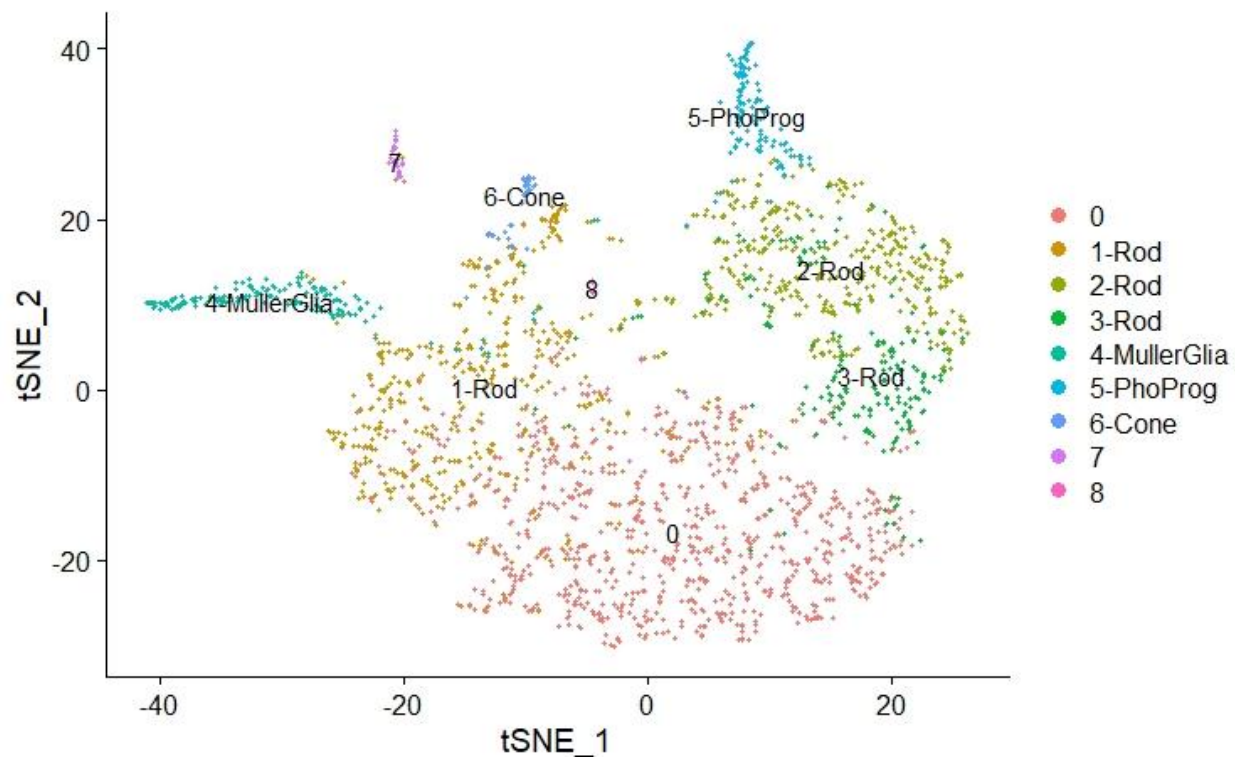


Fig3.3.2.11: WT1 non-rod tSNE plot

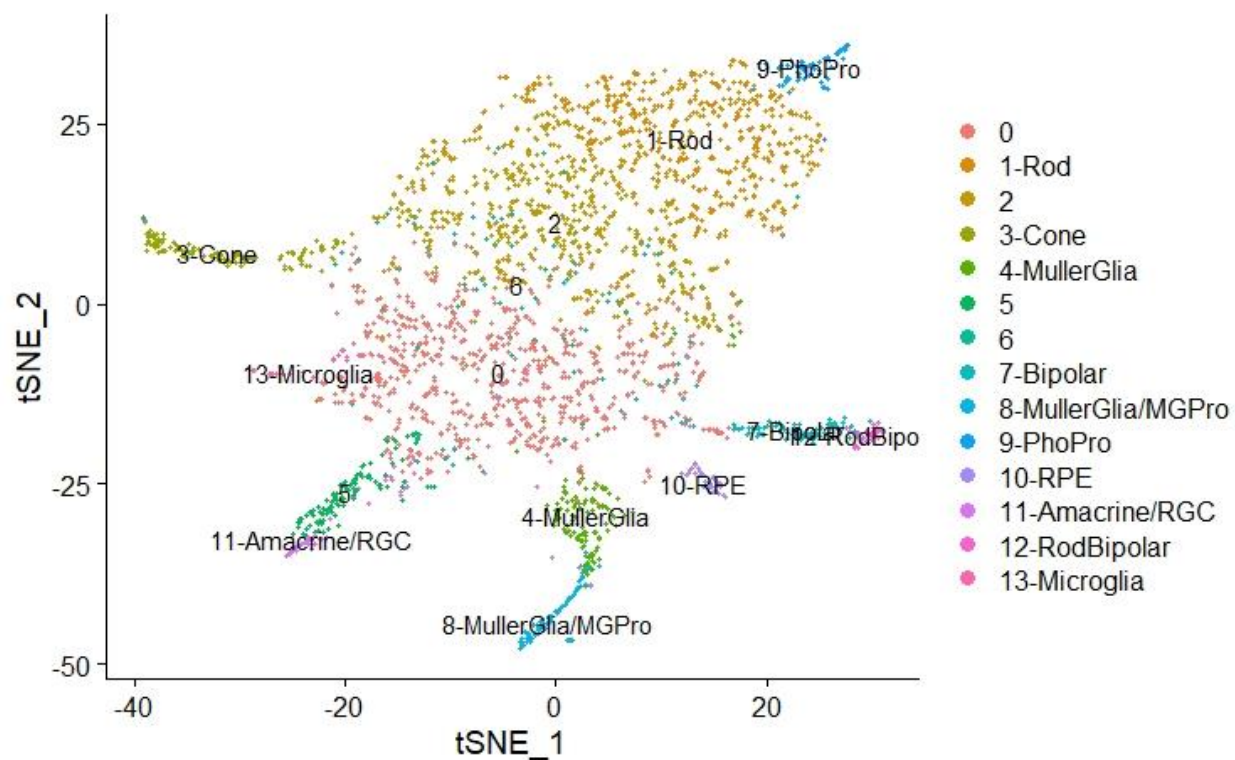


Fig3.3.2.12: WT2 non-rod tSNE plot

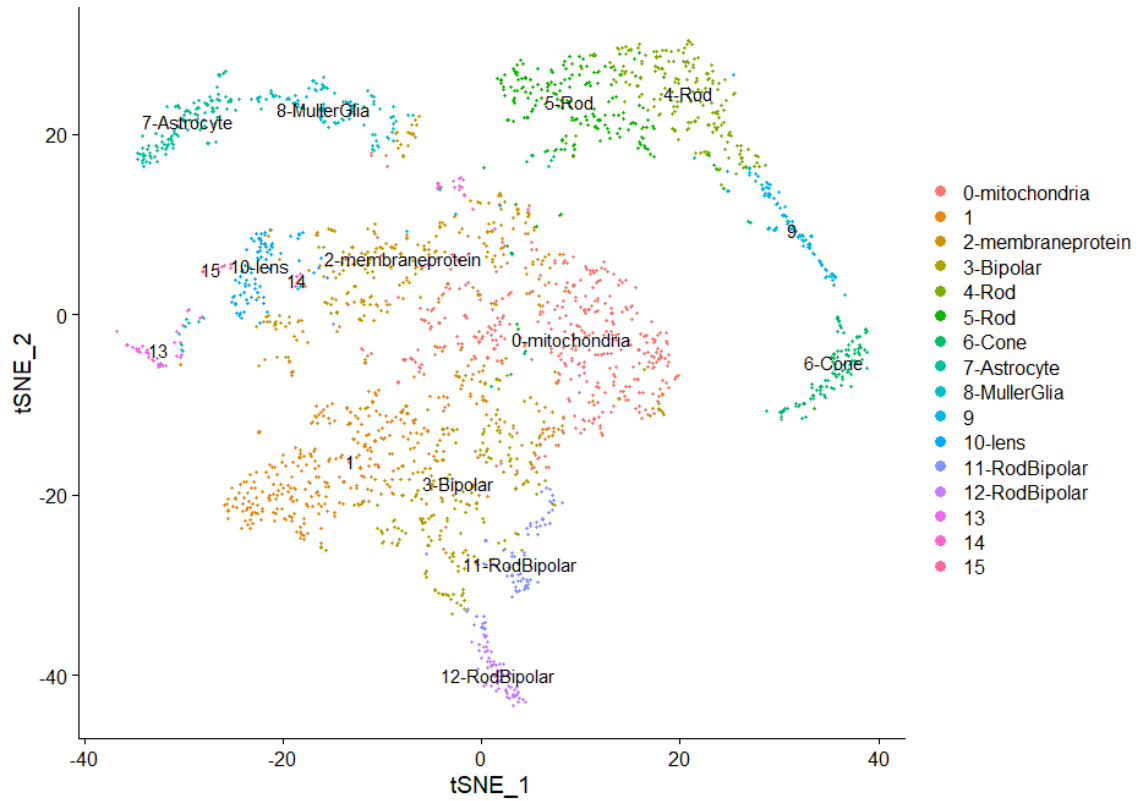


Fig3.3.2.13: WT3 non-rod tSNE plot

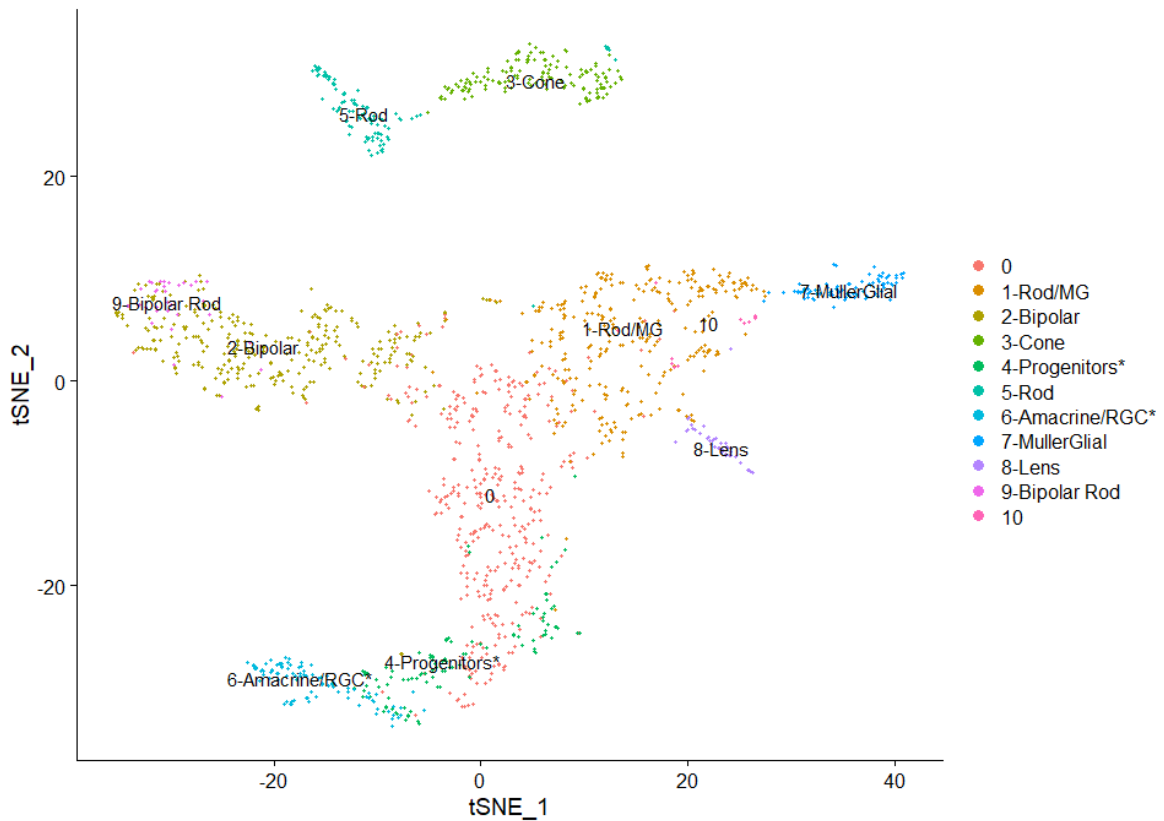


Fig3.3.2.14: KO1 non-rod tSNE plot

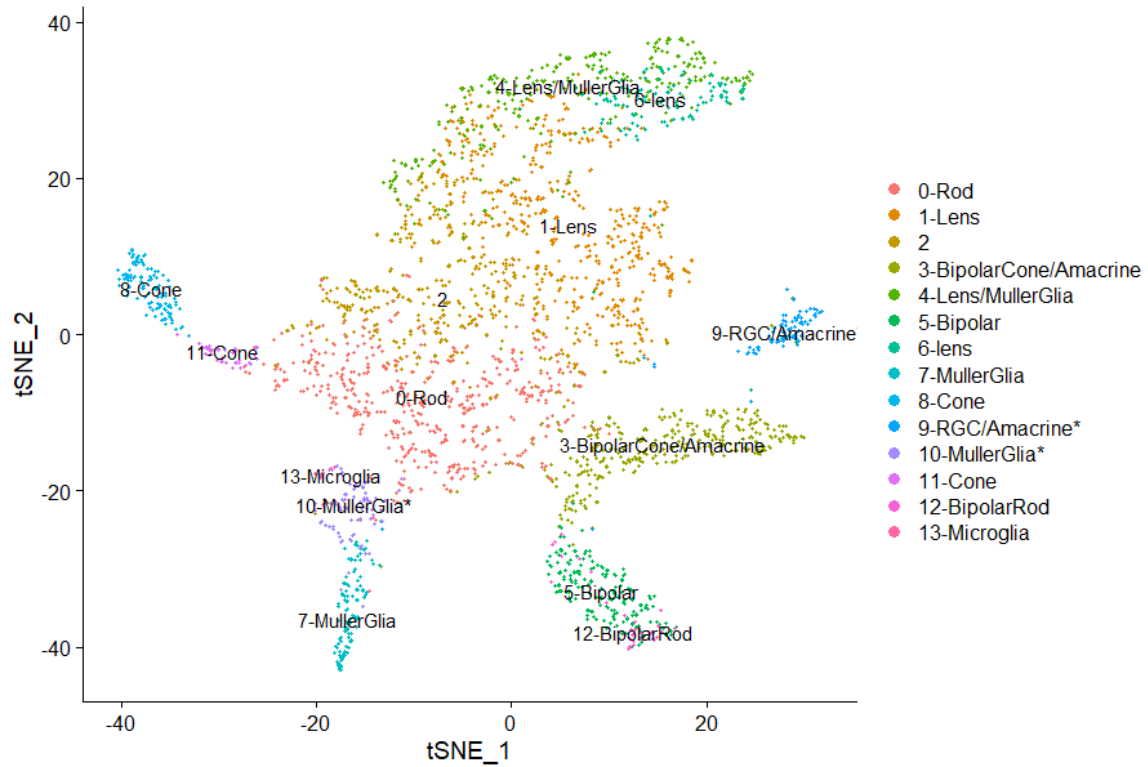


Fig3.3.2.15: KO2 non-rod tSNE plot

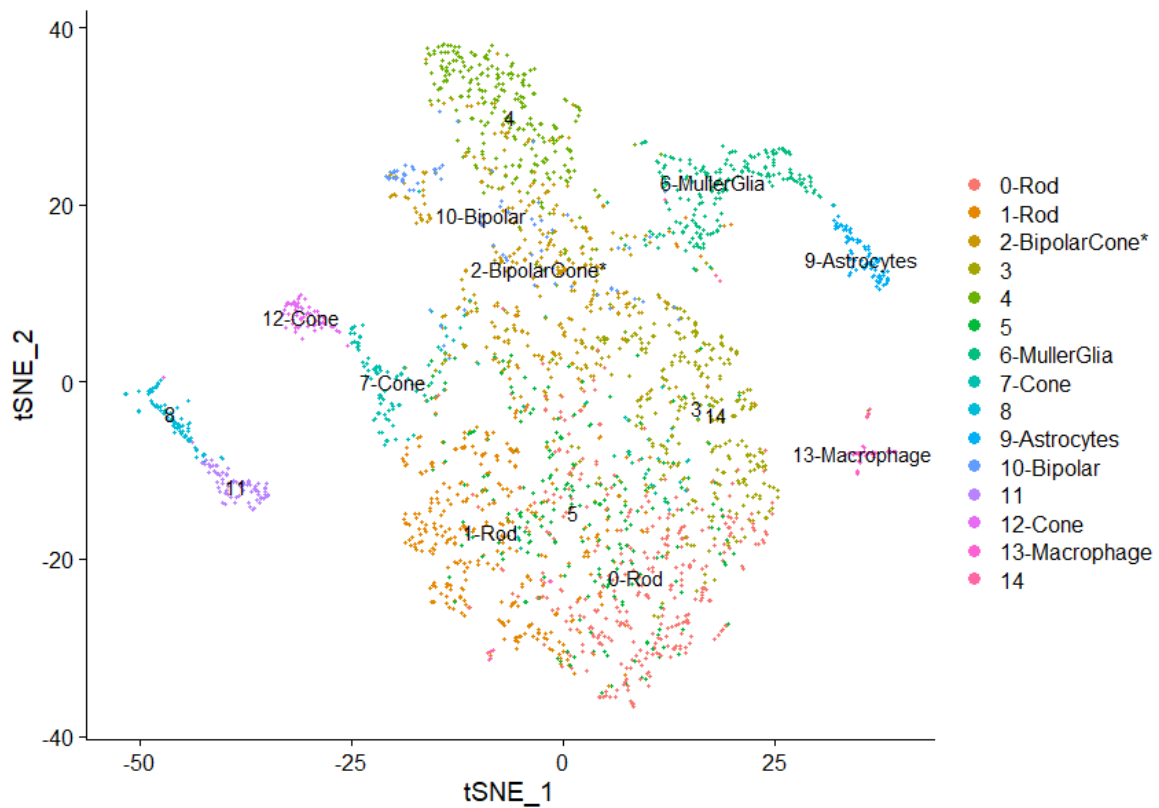


Fig3.3.2.16: KO3 non-rod tSNE plot

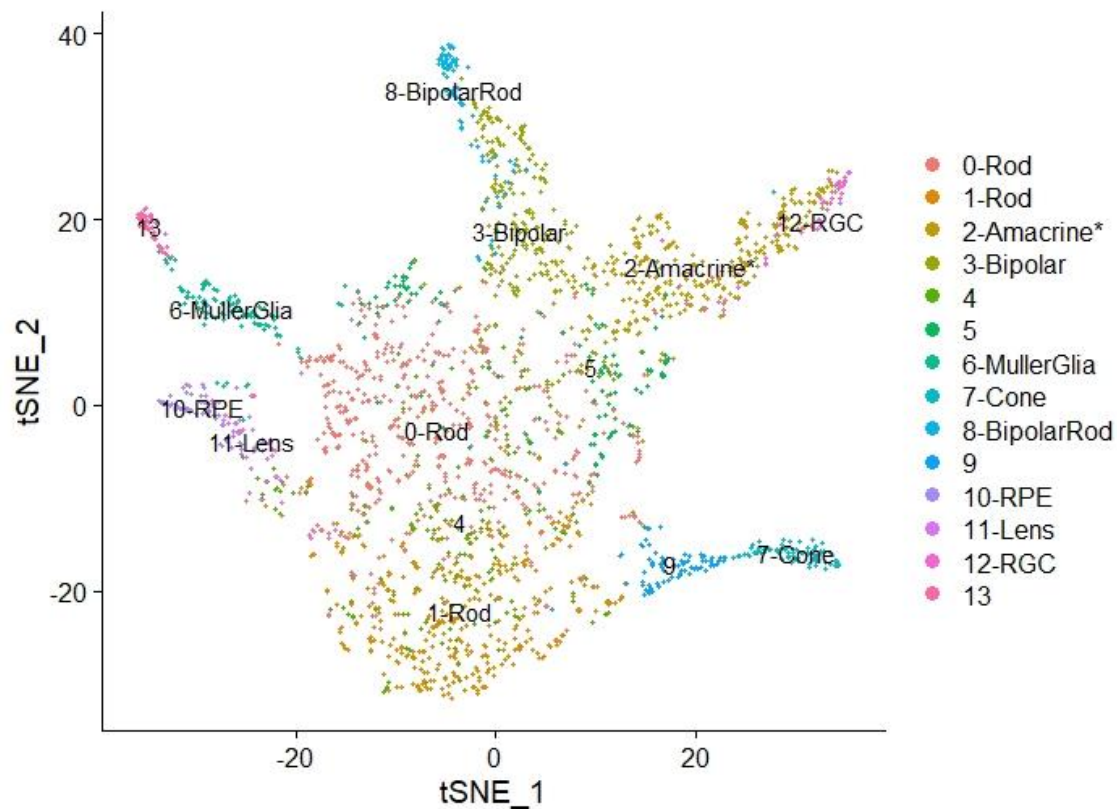


Fig3.3.2.17: KO4 non-rod tSNE plot

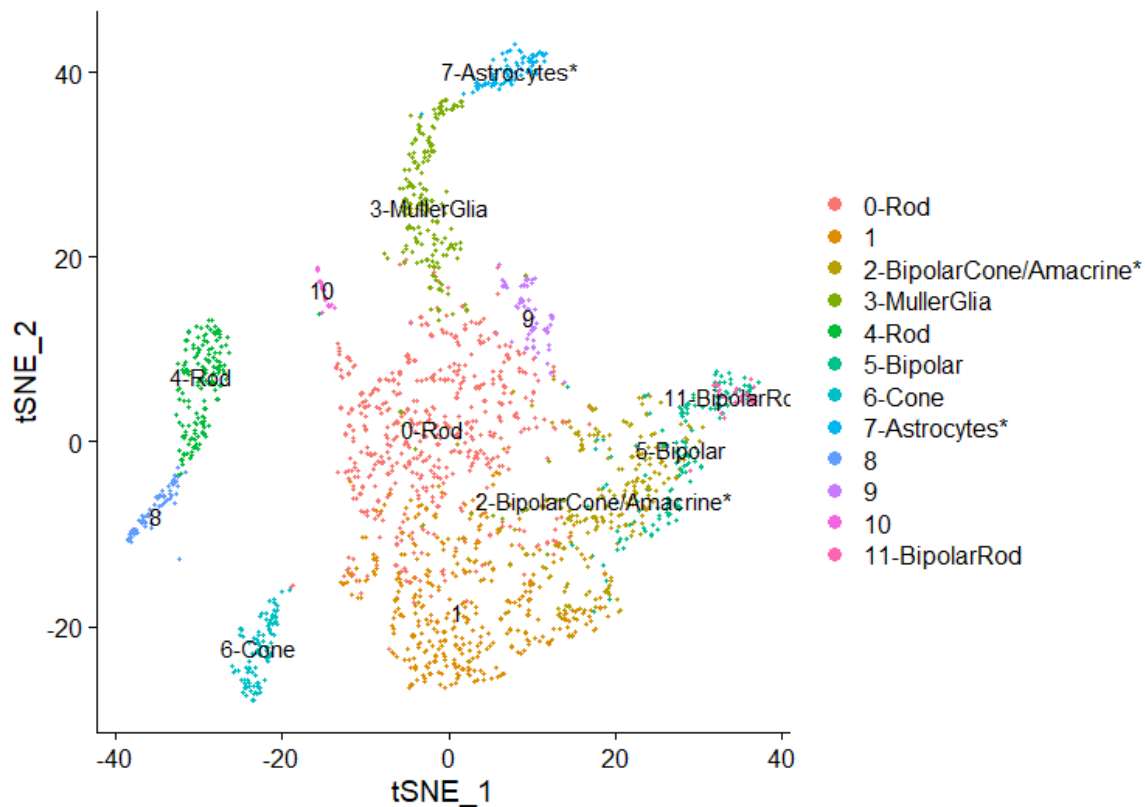


Fig3.3.2.18: KO5 non-rod tSNE plot

3.3.3 Samples Integration

The inspection of individual samples proved the detection and presence of different retinal cell types across all samples. To understand the biological differences between WT and KO samples, the 8 samples were integrated to remove batch effects and technical variations, allowing cell types from across the samples to be projected and clustered together for comparison.

Cell expression matrices output from *CellRanger* for the 8 individual samples were loaded into *Seurat* and analysed using the Integration pipeline as described in the Methods chapter 2.3.2.2. A small filter was applied to remove cells with less than 100 features expressed, which are likely GEMs containing only background contamination reads. Following the pipeline of normalization, integration, scaling, PCA, and UMAP projection, all 62,479 cells combined from the 8 samples were projected onto a UMAP shown in Fig.3.3.3.1

Using the default algorithm in the *FindClusters* functions by *Seurat*, 32 clusters were formed using the top 20 PCs and a granularity setting of 1.5, shown in Fig.3.3.3.3. Number of PCs were chosen based on the Elbow plot (Fig.3.3.3.2), and the granularity was selected such that graphically separated clusters can be well-defined without over segregating large clusters into unnecessary subclusters. Differentially expressed genes of different clusters were identified and labelled with the cell type identity based on high expression of corresponding cell-markers. Expression of cell-type markers of different clusters were summarized in Fig.3.3.3.4, and the expression levels were showcased from Fig.3.3.3.5 to Fig.3.3.3.8. 6 unidentified clusters express some markers of progenitors, amacrine and horizontal cells, but since these cell types were not the focus of this study, they remained as it is for subsequent analysis.

Table 3.3.3.1-2 shows the cell type distribution of the integrated data, again revealing most cells were rod photoreceptors. Distribution of other cell types remained fairly consistent across different samples, suggesting the success of the integration process by clustering similar cell types into graphical proximity, allowing downstream comparison and analysis to be performed.

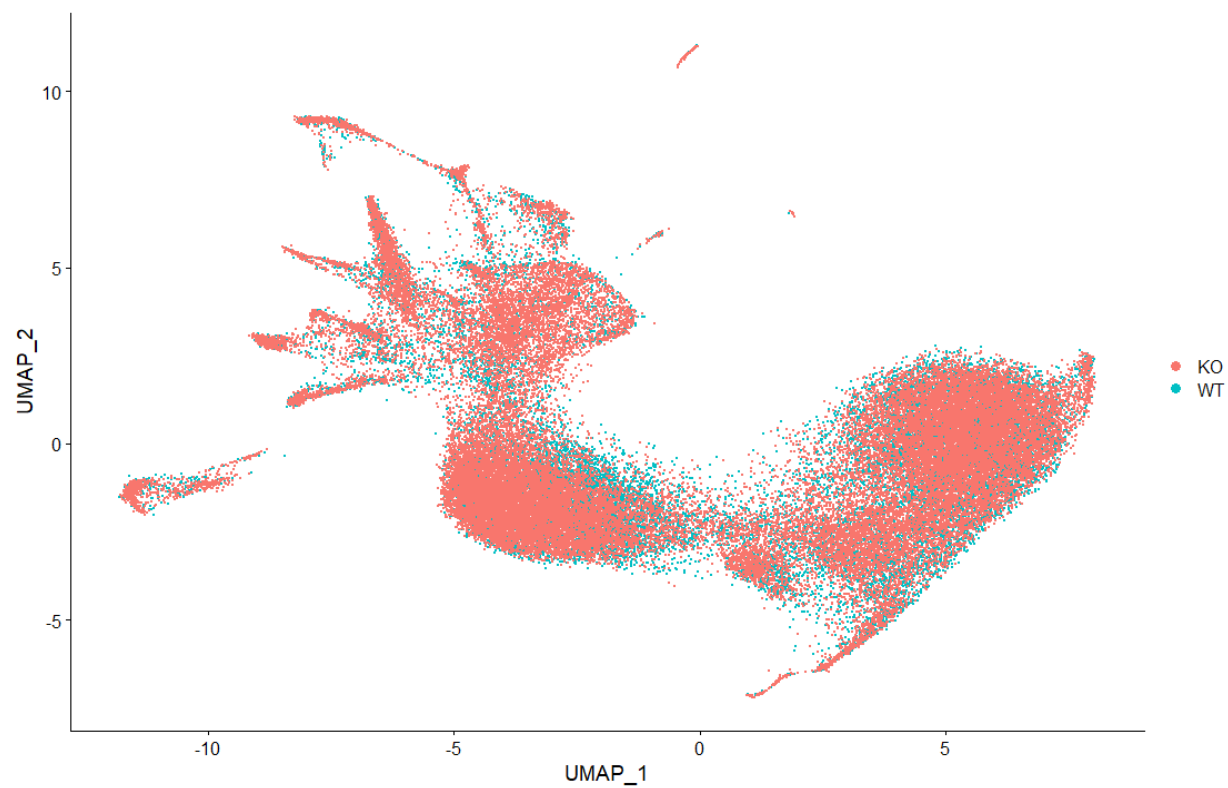


Fig3.3.3.1: UMAP projection of all 62,479 cells from 3 WT and 5 KO samples.

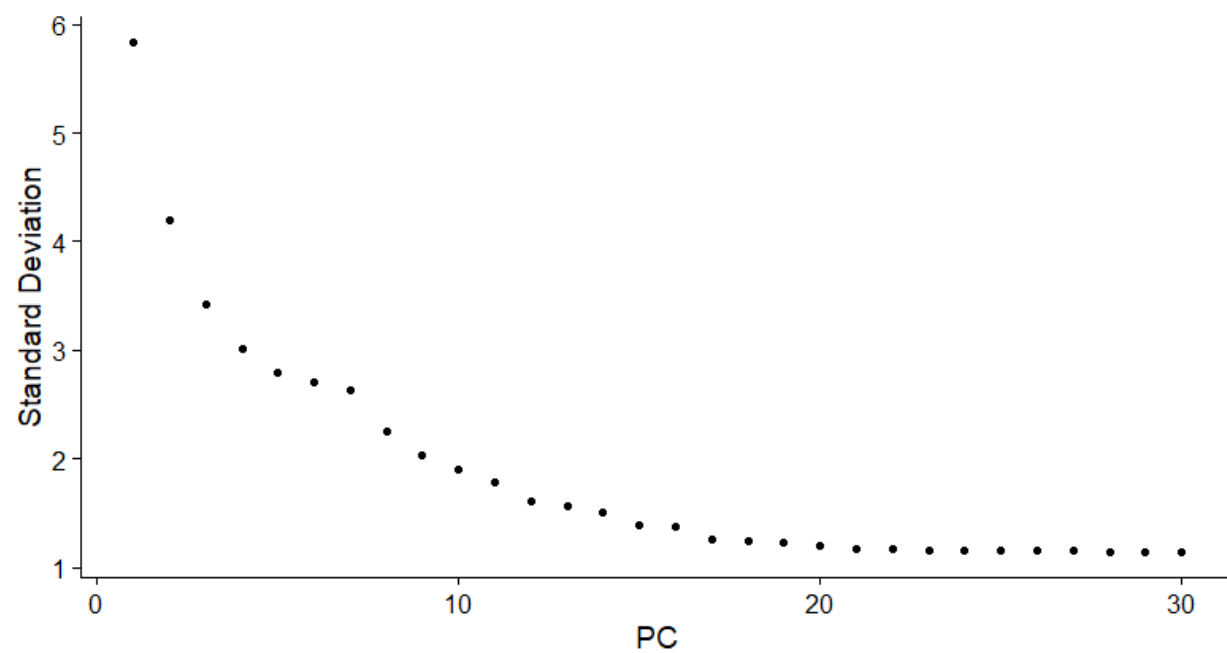


Fig3.3.3.2 Elbow Plot from PCA of the integrated dataset of ABCA1-/- experiment.

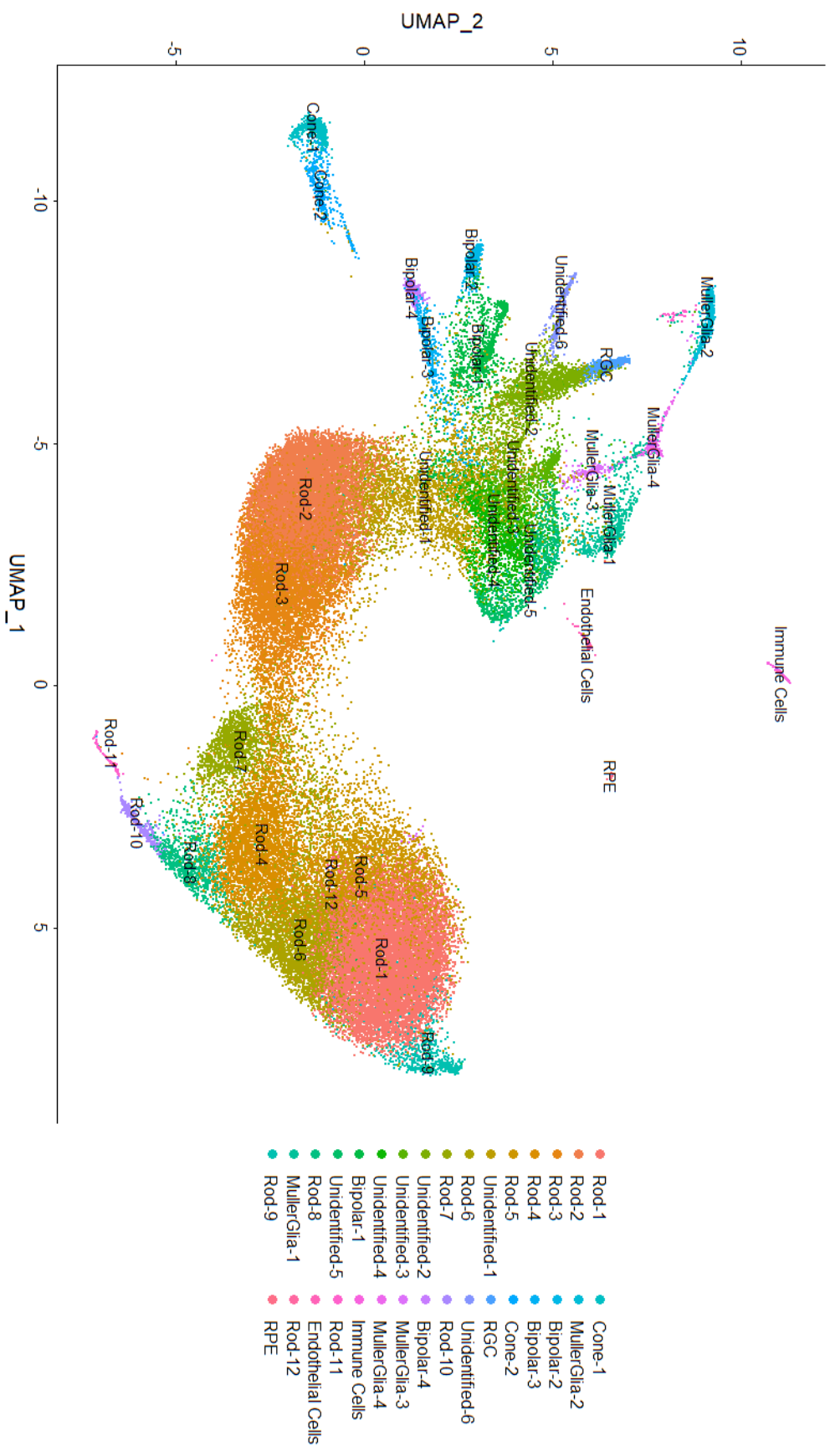


Fig3.3.3.3 UMAP projection of clustered and labelled cells from the integration of all 8 samples

| Sample | No. cells (min 100 features) | No. Rod | No. Cone | No. Müller Glia | No. Astrocyte | No. Bipolar | No. RGC | No. Immun e Cells | No. RPE | No. Endothelial Cells |
|--------|------------------------------------|------------|-------------|-----------------------|------------------|----------------|------------|-------------------------|------------|-----------------------------|
| WT1 | 5526 | 4788 | 39 | 140 | 30 | 56 | 2 | 9 | 2 | 21 |
| WT2 | 6488 | 5087 | 115 | 140 | 59 | 164 | 18 | 20 | 6 | 11 |
| WT3 | 8954 | 6628 | 126 | 217 | 102 | 552 | 38 | 11 | 8 | 15 |
| KO1 | 6890 | 5186 | 169 | 151 | 76 | 488 | 14 | 12 | 0 | 9 |
| KO2 | 8144 | 5235 | 149 | 113 | 72 | 358 | 32 | 10 | 0 | 12 |
| KO3 | 8683 | 6151 | 155 | 171 | 80 | 267 | 15 | 33 | 0 | 8 |
| KO4 | 8969 | 6670 | 169 | 175 | 62 | 350 | 31 | 23 | 11 | 11 |
| KO5 | 8825 | 6753 | 119 | 224 | 83 | 255 | 6 | 25 | 3 | 7 |

Table 3.3.3.1 Cell number counts for different cell types.

| Sample | % Rod | % Cone | % MG | % Astrocyte | % Bipolar | % RGC | % Immune Cells | % RPE | % Endothelial Cells |
|--------|--------|--------|-------|-------------|-----------|-------|----------------|-------|---------------------|
| WT1 | 86.645 | 0.706 | 2.533 | 0.543 | 1.013 | 0.036 | 0.163 | 0.036 | 0.380 |
| WT2 | 78.406 | 1.773 | 2.158 | 0.909 | 2.528 | 0.277 | 0.308 | 0.092 | 0.170 |
| WT3 | 74.023 | 1.407 | 2.423 | 1.139 | 6.165 | 0.424 | 0.123 | 0.089 | 0.168 |
| KO1 | 75.269 | 2.453 | 2.192 | 1.103 | 7.083 | 0.203 | 0.174 | 0.000 | 0.131 |
| KO2 | 64.280 | 1.830 | 1.388 | 0.884 | 4.396 | 0.393 | 0.123 | 0.000 | 0.147 |
| KO3 | 70.840 | 1.785 | 1.969 | 0.921 | 3.075 | 0.173 | 0.380 | 0.000 | 0.092 |
| KO4 | 74.367 | 1.884 | 1.951 | 0.691 | 3.902 | 0.346 | 0.256 | 0.123 | 0.123 |
| KO5 | 76.521 | 1.348 | 2.538 | 0.941 | 2.890 | 0.068 | 0.283 | 0.034 | 0.079 |

Table3.3.3.2 Cell counts in percentage for different cell types.

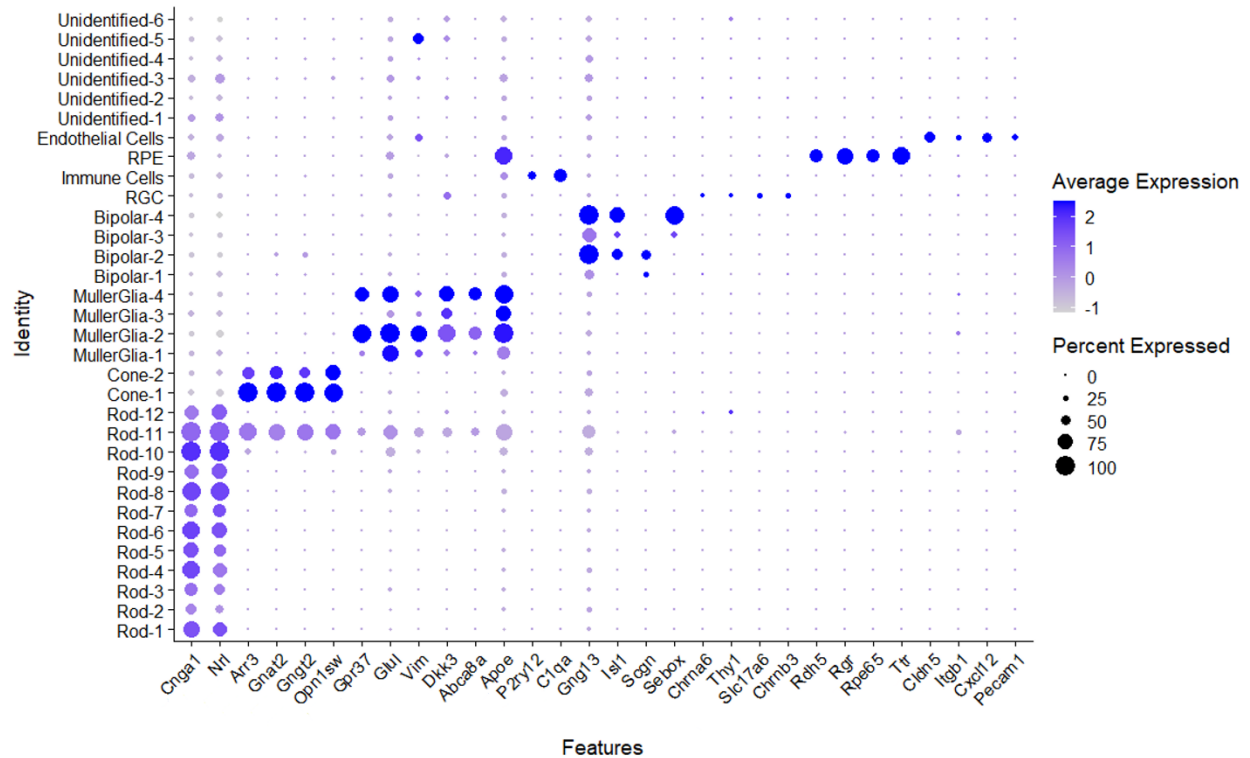


Fig3.3.3.4 Dot Plot of the integrated samples, displaying expression level of different cell-type markers by different clusters.

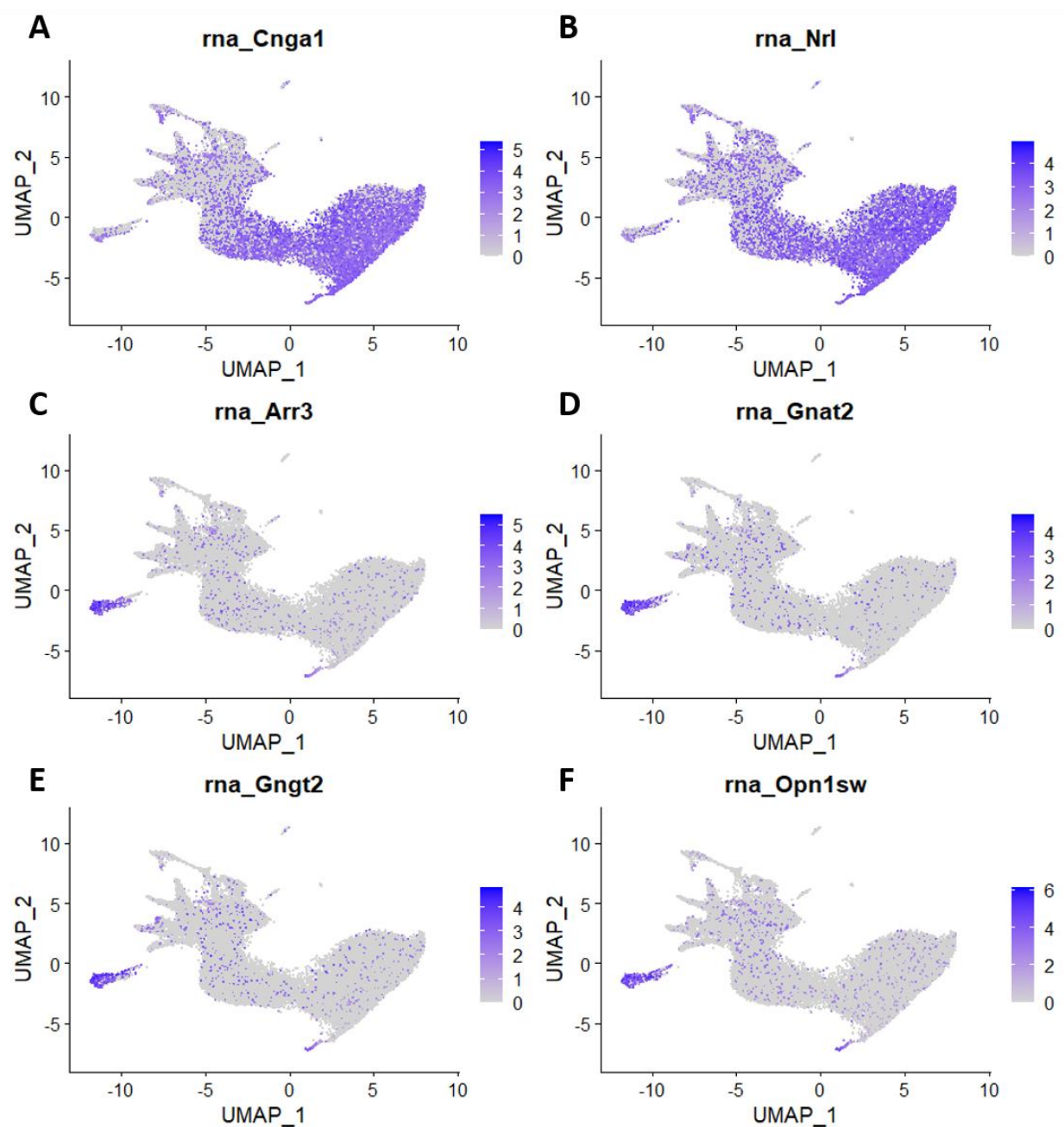


Fig3.3.3.5A-B: Rod cells markers expression. C-F: Cone cell markers expression.

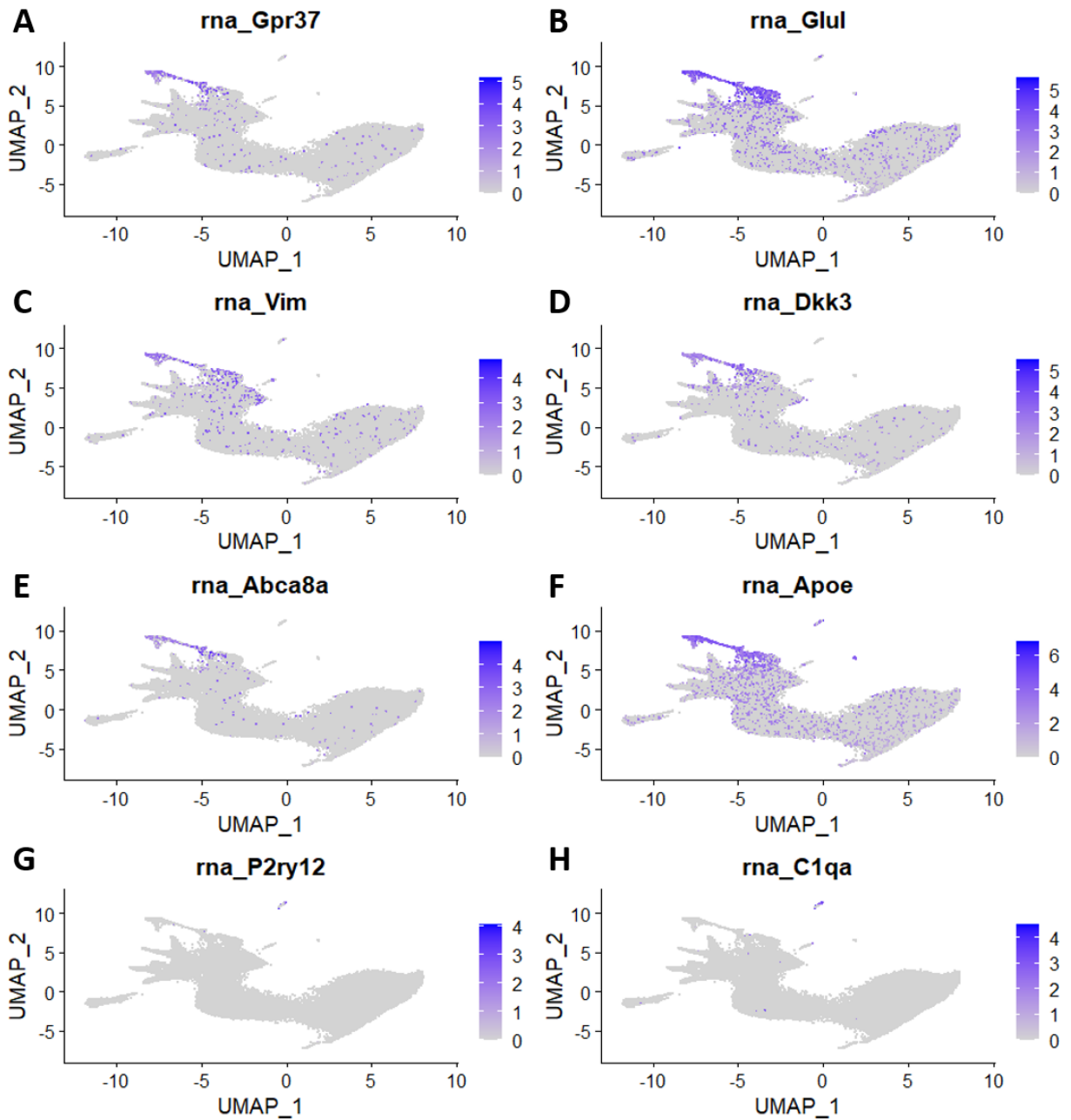


Fig3.3.3.6A-F: Müller Glia cell/ astrocyte markers expression. G-H: Microglia cell markers expression

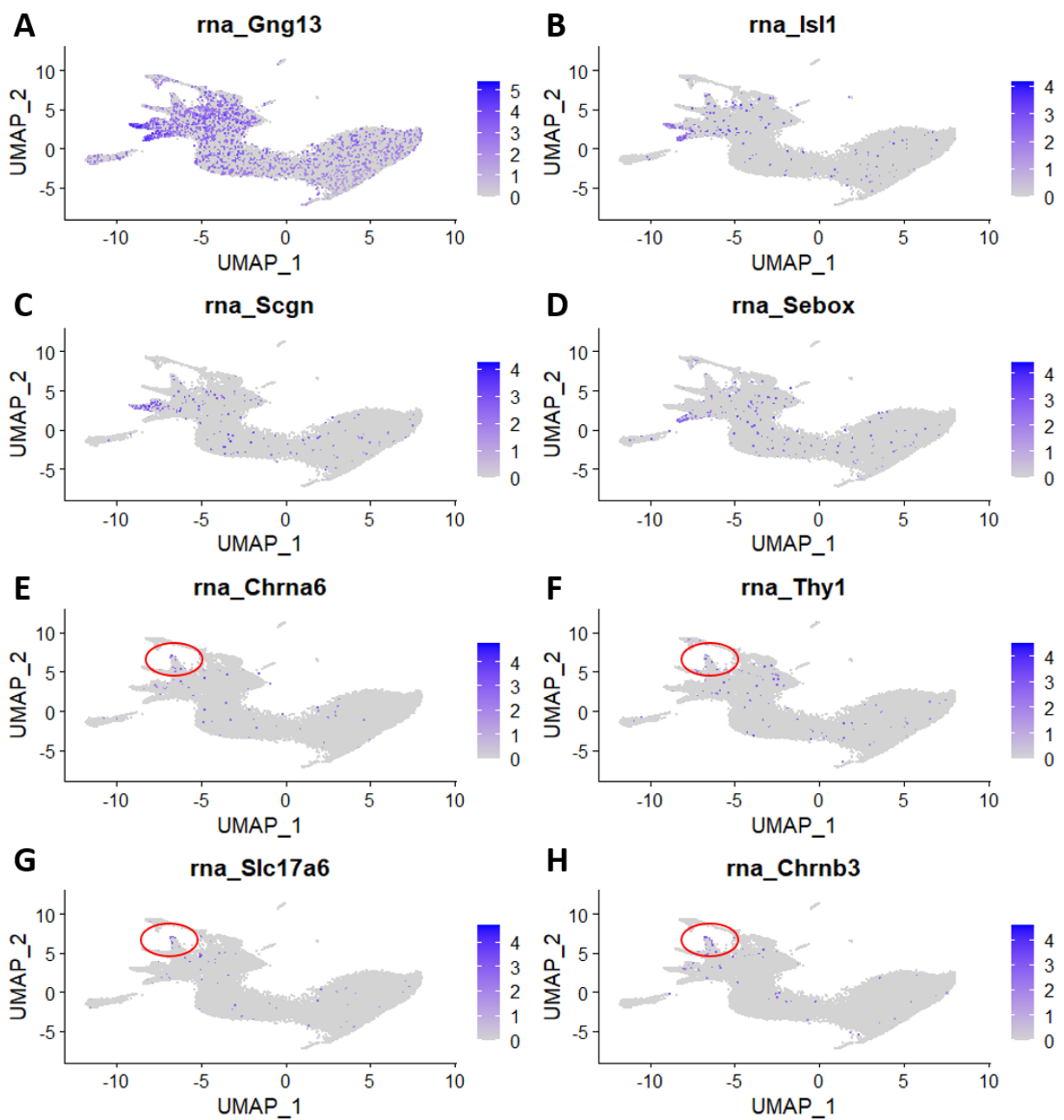


Fig3.3.3.7A-D: Bipolar cell markers expression. E-H: RGC markers expression

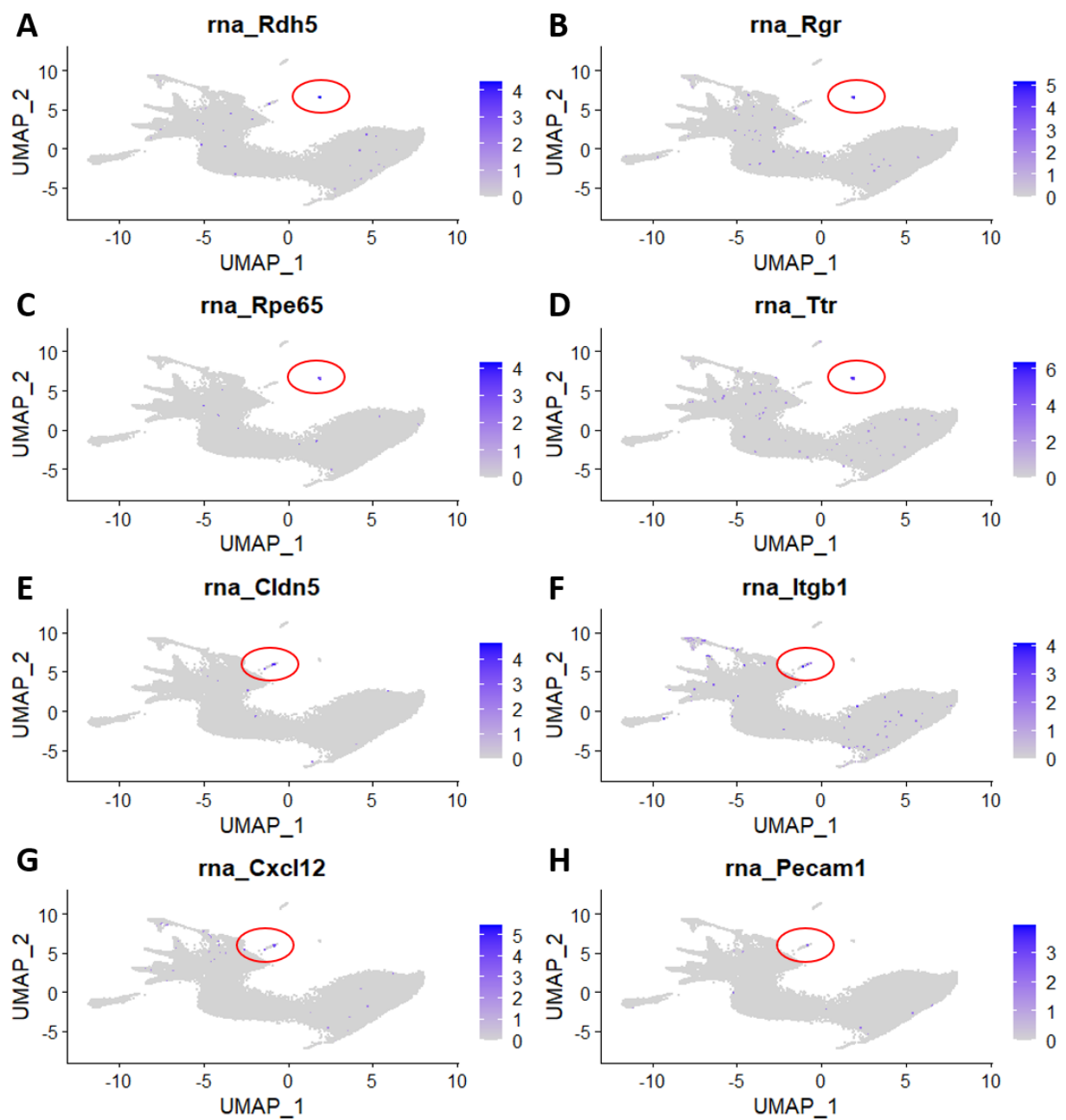


Fig3.3.3.8A-D: RPE cell markers expression. E-H: Endothelial cell markers expression

3.3.4 WT vs KO Ontological analysis

The initial result of the integration pipeline showed success in removing batch effects between different samples and allowed cells from both WT and KO samples to be projected into proximity of their respective cell types. The differences between the adjusted expression profiles of WT and KO cells shall reflect mainly biological differences, making them more comparable. In order to gain initial insight into the biological differences, gene expression across all cells in WT samples were averaged and compared against that of KO samples. The differential expression gene list of WT vs KO samples was then analysed using Ingenuity Pathway Analysis (IPA) to perform ontological analysis. Table3.3.4.1 shows the 50 topmost significantly affected Canonical Pathways in KO samples in comparison to WT, ranked by $-\log(p\text{-value})$.

Among the most significantly affected pathways, remarkably many of them are related to neuroinflammation, such as “IL-8 Signaling”, “CXCR4 Signaling”, “CCR5 Signaling in Macrophages” and “Neuroinflammation Signaling Pathway”. These pathways all have a positive z-score, signifying their activation in KO samples. Fig.3.3.4.1 shows the pathway diagram for “Neuroinflammation Signaling Pathway”, with multiple downstream mechanisms predicted to be activated, including “Oxidative stress”, “Astrogliosis”, “Neurons damage”, “T cell recruitment” and “Microglia activation”. Multiple pro-inflammatory proteins such as members of the chemokine family were shown to be significantly expressed, signifying a strong inflammatory and immune response. Other neuroinflammation related pathways such as “Chemokine Signaling”, “TGF- β Signaling” are compiled in Table3.3.4.2, all showing significant activation. Thus, evidence of neuroinflammation found in the average expression of cells in retinal samples of ABCA^{-/-} were very strong.

| Ingenuity Canonical Pathways | -log(p-value) | z-score |
|-----------------------------------------------------|---------------|---------|
| Osteoarthritis Pathway | 13.6 | -0.447 |
| RhoGDI Signaling | 13.3 | -2.611 |
| Phototransduction Pathway | 13.2 | NaN |
| Hepatic Fibrosis / Hepatic Stellate Cell Activation | 12.1 | NaN |
| Signaling by Rho Family GTPases | 12.1 | 3.683 |
| Hepatic Fibrosis Signaling Pathway | 11.7 | 2.216 |
| EIF2 Signaling | 11.3 | -1.347 |
| Androgen Signaling | 10.3 | 1.606 |
| CREB Signaling in Neurons | 9.82 | 2.449 |
| IL-8 Signaling | 9.78 | 2.846 |
| Opioid Signaling Pathway | 9.74 | 2.48 |
| Cardiac Hypertrophy Signaling | 9.68 | 1.761 |
| Relaxin Signaling | 9.09 | 1.414 |
| CXCR4 Signaling | 8.89 | 1.512 |
| Estrogen Receptor Signaling | 8.46 | 1.664 |
| Synaptogenesis Signaling Pathway | 8.4 | 3.221 |
| Role of NFAT in Cardiac Hypertrophy | 8.3 | 1.947 |
| Ephrin B Signaling | 8.25 | 1.604 |
| Tec Kinase Signaling | 8 | 1.964 |
| Axonal Guidance Signaling | 7.99 | NaN |
| Glutamate Receptor Signaling | 7.91 | 2.449 |
| G Beta Gamma Signaling | 7.84 | 1.732 |
| Calcium Signaling | 7.8 | 3.024 |
| Thrombin Signaling | 7.68 | 2.558 |
| mTOR Signaling | 7.56 | 1.528 |
| Ephrin Receptor Signaling | 7.44 | 1.768 |
| Protein Kinase A Signaling | 7.34 | -0.146 |
| White Adipose Tissue Browning Pathway | 7.27 | 2.646 |
| CCR5 Signaling in Macrophages | 7.24 | 1.667 |
| ILK Signaling | 6.82 | 1.915 |
| Epithelial Adherens Junction Signaling | 6.7 | NaN |
| Corticotropin Releasing Hormone Signaling | 6.65 | 1.177 |
| Role of NFAT in Regulation of the Immune Response | 6.38 | 1.177 |
| GNRH Signaling | 6.37 | 2.6 |
| GABA Receptor Signaling | 5.9 | NaN |
| Gαq Signaling | 5.81 | 1.569 |
| GP6 Signaling Pathway | 5.78 | 3 |
| eNOS Signaling | 5.75 | 1.342 |
| Endocannabinoid Neuronal Synapse Pathway | 5.69 | 2.449 |
| Phospholipase C Signaling | 5.6 | 1.219 |
| Germ Cell-Sertoli Cell Junction Signaling | 5.54 | NaN |
| Neuroinflammation Signaling Pathway | 5.34 | 1.372 |
| cAMP-mediated signaling | 5.29 | 1.859 |
| Cardiac Hypertrophy Signaling (Enhanced) | 5.27 | 3.302 |
| Human Embryonic Stem Cell Pluripotency | 5.24 | NaN |
| Molecular Mechanisms of Cancer | 5.17 | NaN |
| G Protein Signaling Mediated by Tubby | 5.13 | NaN |
| Reelin Signaling in Neurons | 5.12 | 2.041 |
| Leukocyte Extravasation Signaling | 5.08 | 1.461 |
| IL-1 Signaling | 5.06 | 1.633 |

Table3.3.4.1 Top 50 most significantly affected Canonical Pathways in differential expression between all cells from KO vs WT samples. Yellow and green highlights label pathways related to neuroinflammation and senescence respectively.

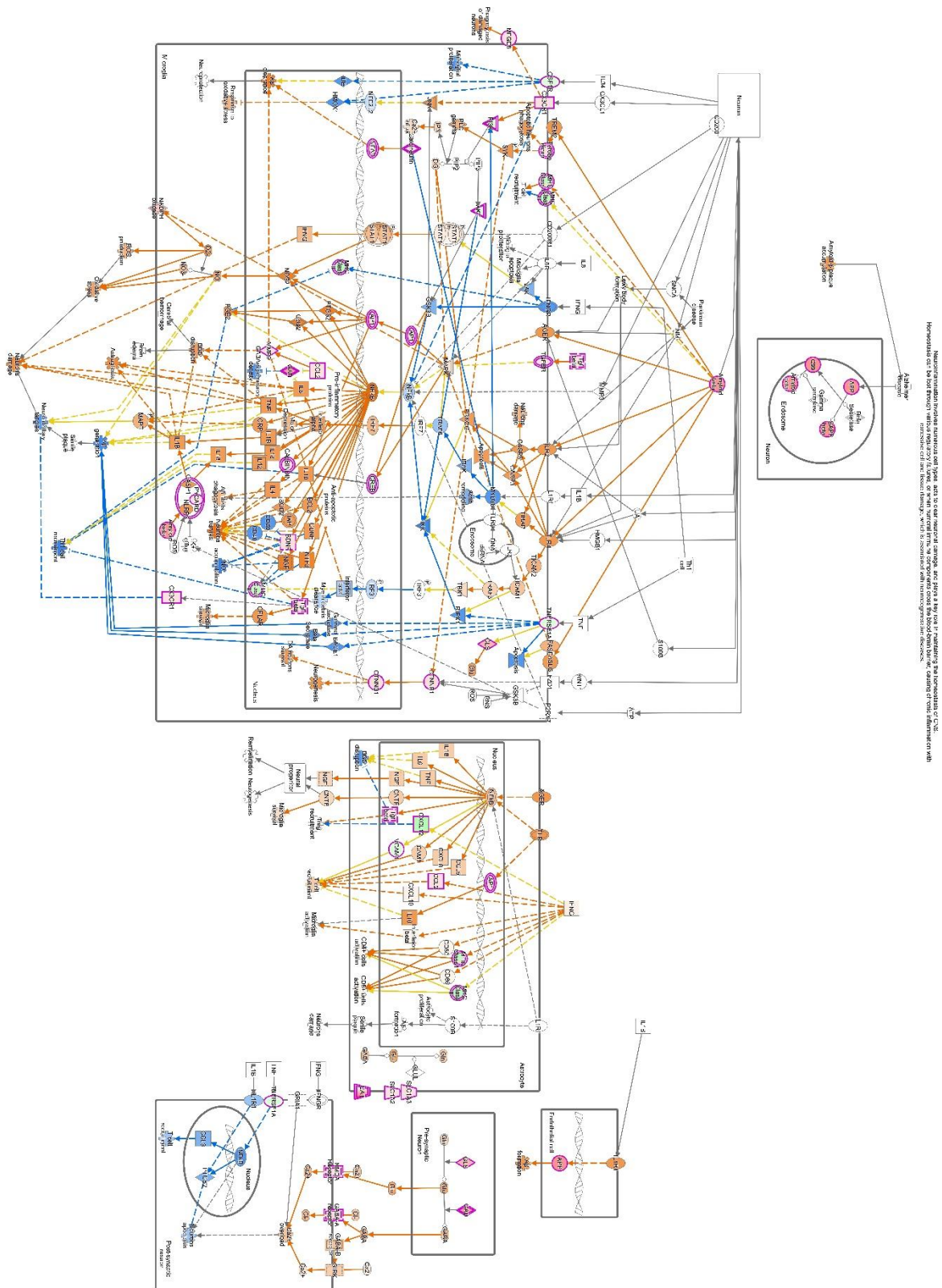


Fig.3.3.4.1 Neuroinflammation Signalling pathway diagram from IPA, with predicted activation of mechanisms such as “microglia activation”, “astrogliosis”, “neurons damage” observed in KO samples.

| Ingenuity Canonical Pathways | -log(p-value) | z-score |
|---------------------------------------------------|---------------|---------|
| IL-8 Signaling | 9.78 | 2.846 |
| CXCR4 Signaling | 8.89 | 1.512 |
| Role of NFAT in Regulation of the Immune Response | 6.38 | 1.177 |
| Neuroinflammation Signaling Pathway | 5.34 | 1.372 |
| Leukocyte Extravasation Signaling | 5.08 | 1.461 |
| IL-1 Signaling | 5.06 | 1.633 |
| Chemokine Signaling | 3.66 | 1.5 |
| PKC θ , Signaling in T Lymphocytes | 2.28 | 2.138 |
| PI3K Signaling in B Lymphocytes | 1.87 | 1.5 |
| Interferon Signaling | 1.82 | 0.378 |
| Notch Signaling | 1.76 | 0.447 |
| IL-6 Signaling | 1.65 | 0.775 |
| TGF- β Signaling | 1.61 | 1.508 |
| IL-3 Signaling | 1.51 | 1.508 |
| IL-15 Production | 1.47 | 1.291 |
| NF- κ B Activation by Viruses | 1.41 | 0.302 |

Table3.3.4.2 Neuroinflammation related pathways activated in KO vs WT all cells differential expression.

Table3.3.4.1 also labelled multiple Canonical Pathways with potential relation to cellular senescence being significantly activated. Other significantly activated senescence related pathways are shown in Table 3.3.4.3. In particular, “mTOR signaling” was shown to be significantly activated with a -log(p-value) of 7.56 and z-score of 1.528. This pathway had been shown to be highly associated with multiple age-related diseases including neurodegeneration, as well as senescence. Senescent cells were known to secrete pro-inflammatory molecules through the process known as senescence-associated secretory phenotype (SASP), with related pathways are shown in Table3.3.4.4. There are multiple overlaps between neuroinflammation and SASP related pathways, thus this data suggests the potential mechanism of senescence-induced neuroinflammation in KO retinal tissues, causing age-dependent phenotype of retinal degeneration, which could contribute to our understanding of the pathogenesis of glaucoma.

| Ingenuity Canonical Pathways | -log(p-value) | z-score |
|------------------------------|---------------|---------|
| Androgen Signaling | 10.3 | 1.606 |
| Estrogen Receptor Signaling | 8.46 | 1.664 |
| Calcium Signaling | 7.8 | 3.024 |
| Thrombin Signaling | 7.68 | 2.558 |
| mTOR Signaling | 7.56 | 1.528 |
| cAMP-mediated signaling | 5.29 | 1.859 |
| p53 Signaling | 3.09 | 0.277 |
| AMPK Signaling | 2.94 | 0.655 |
| BEX2 Signaling Pathway | 2.74 | 1.604 |
| HMGB1 Signaling | 2.28 | 1.964 |
| NF-κB Signaling | 1.61 | 1.091 |
| Senescence Pathway | 1.27 | 0.962 |

Table3.3.4.3 Senescence related pathways activated in KO vs WT all cells differential expression.

| Ingenuity Canonical Pathways | -log(p-value) | z-score |
|-------------------------------|---------------|---------|
| IL-8 Signaling | 9.78 | 2.846 |
| CXCR4 Signaling | 8.89 | 1.512 |
| IL-3 Signaling | 1.51 | 1.508 |
| CCR3 Signaling in Eosinophils | 4.44 | 1.414 |
| IL-1 Signaling | 5.06 | 1.633 |
| BMP signaling pathway | 1.31 | 0.632 |
| HGF Signaling | 1.46 | 0.577 |
| eNOS Signaling | 5.75 | 1.342 |
| VEGF Signaling | 2.22 | 0.832 |

Table3.3.4.4 SASP related pathways activated in KO vs WT all cells differential expression.

3.3.5 Müller Glia and Astrocyte subclusters

3.3.5.1 Identifying subclusters

Müller glia and astrocytes are both classified as macroglia in the retina. Their feature and functions are quite similar, thus they share similar expression profile as well as overlapping cell-type markers. Since the focus of our studies placed emphasis on astrocytes, distinguishing between Müller glia and astrocytes is necessary.

Although 4 clusters in Fig3.3.3.2 were labelled as “Müller Glia” cell clusters, some of the cell markers were common among both cell types. To further investigate and label those cells, cells from the 4 clusters (Müller glia – 1, Müller glia – 2, Müller glia – 3 and Müller glia – 4) were isolated to create a subset of the dataset with potential Müller glia and astrocytes (MG/A cluster). By applying the *Seurat* pipeline as mentioned in the Methods chapter 2.3.2.3, the MG/A subset was clustered into 9 subclusters using a granularity of 0.8, as shown in Fig.3.3.5.1.1.

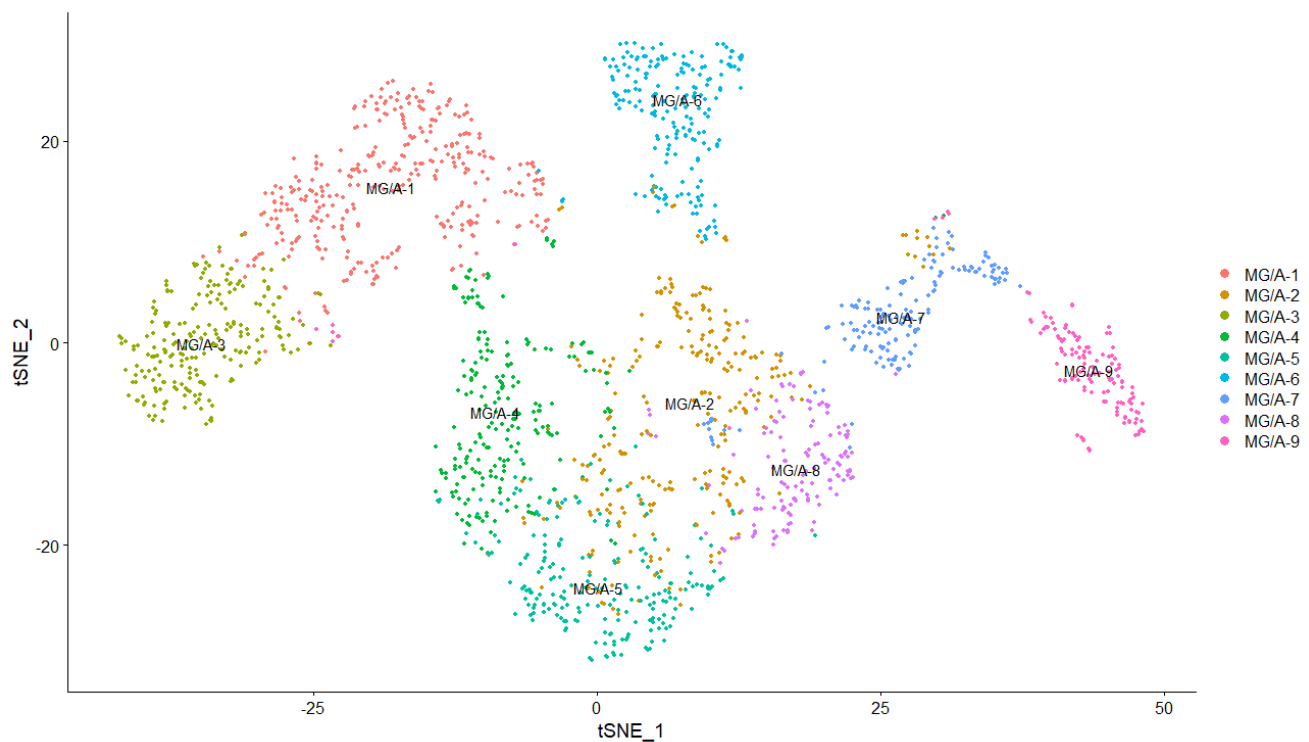


Fig3.3.5.1.1 tSNE plot of the MG/A subcluster

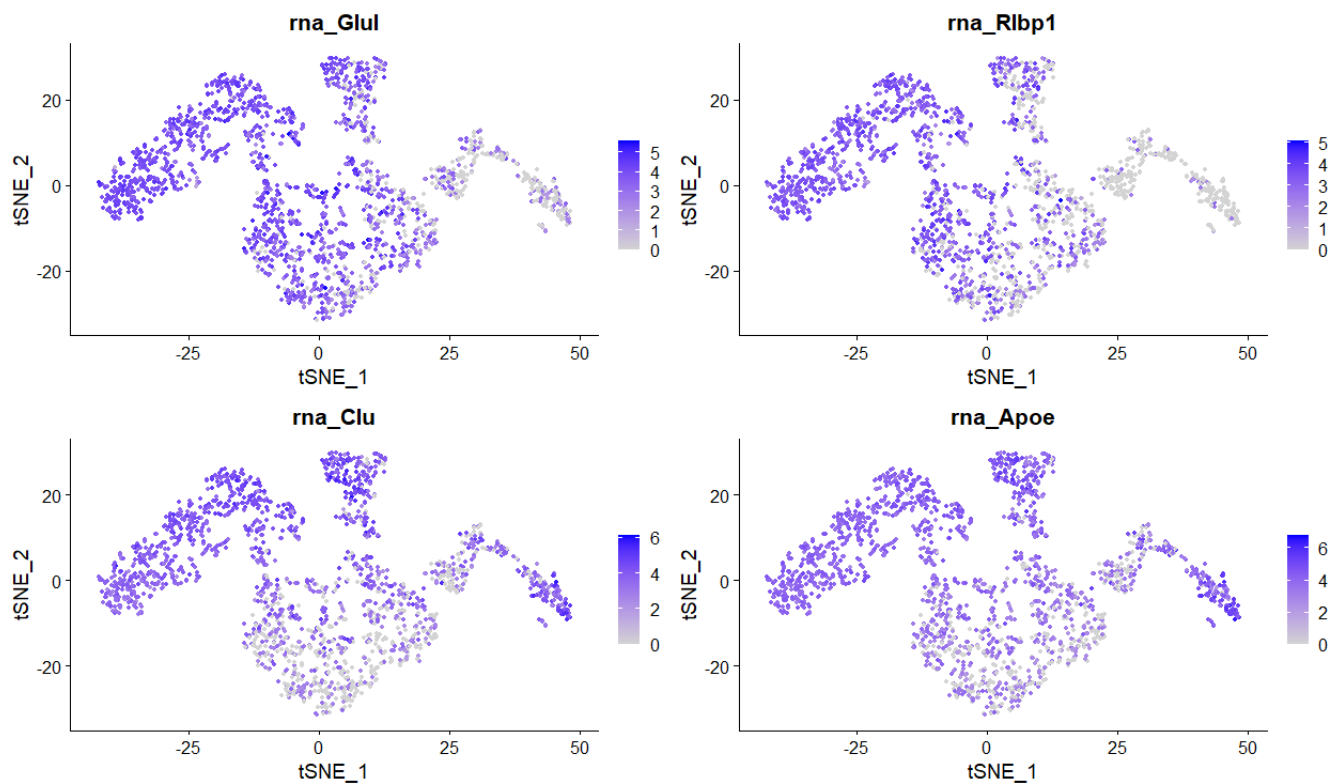


Fig3.3.5.1.2 Müller Glia markers in the MG/A subcluster

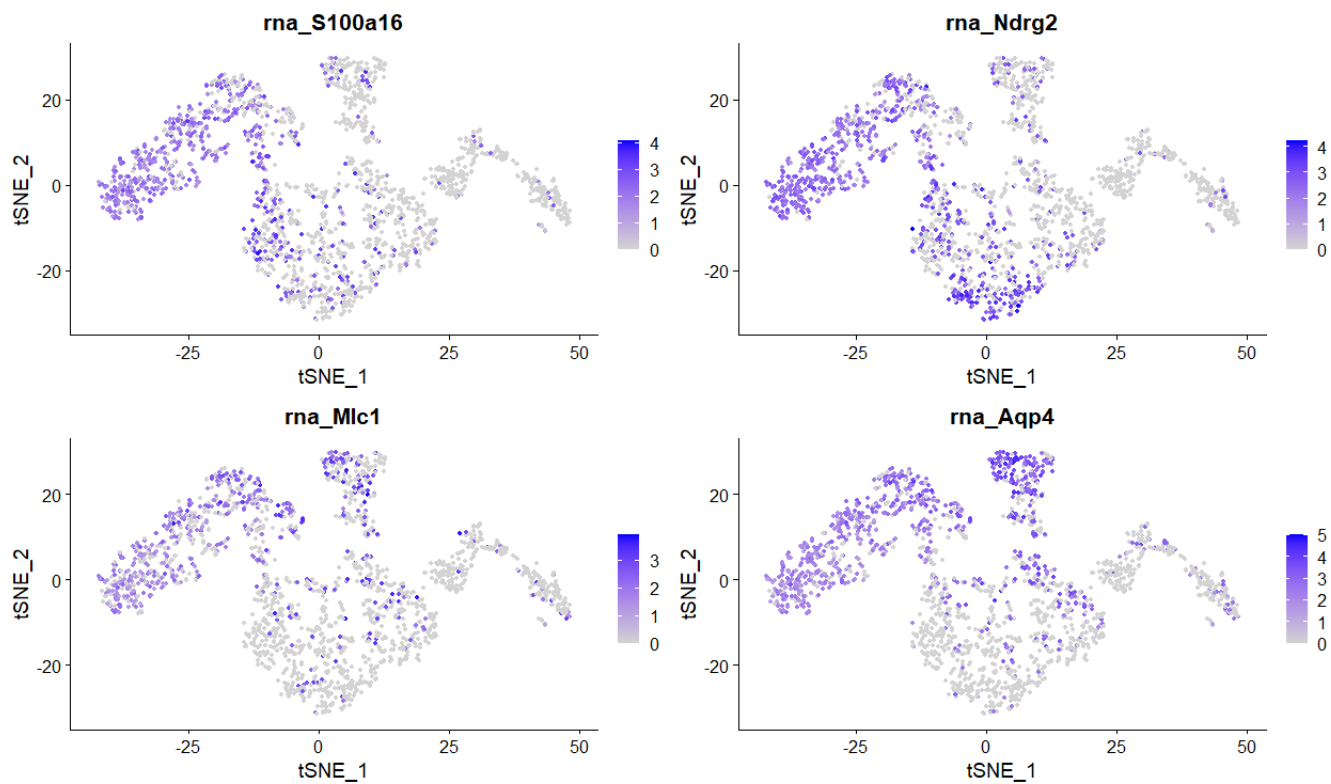


Fig3.3.5.1.3 Astrocytes markers in the MG/A subcluster tSNE plot

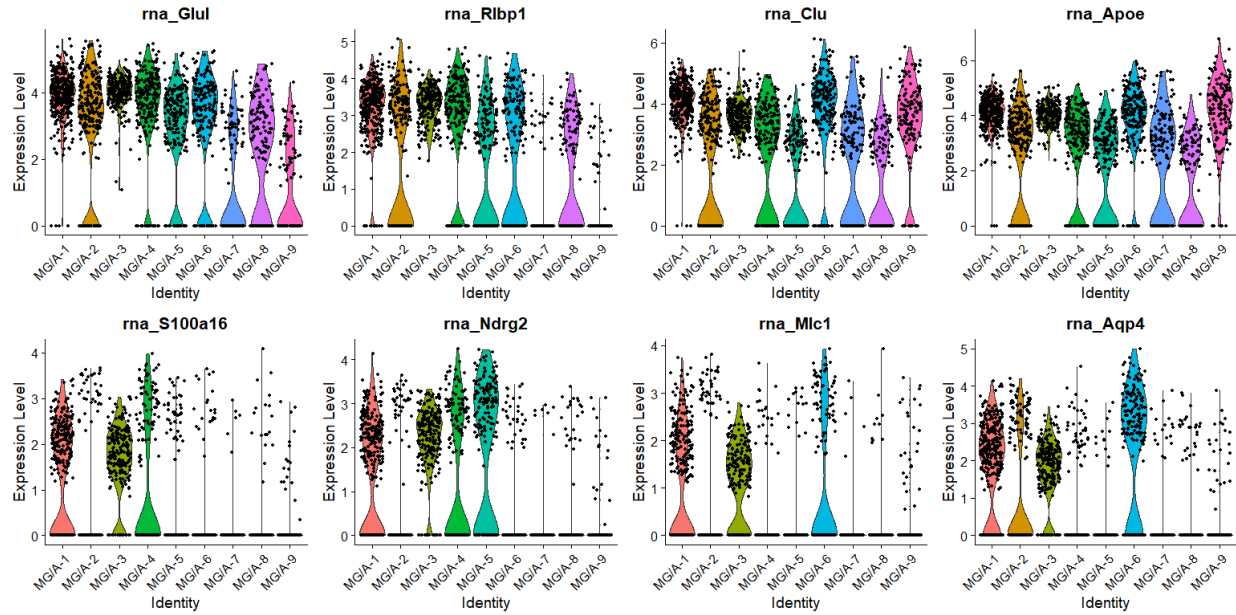


Fig3.3.5.1.4 Violin Plot of Müller glia markers (upper row) and Astrocyte markers (bottom row) expression in MG/A subclusters.

The visualization of MG/A subclusters in Fig3.3.5.1.1 showed distinct segregated regions of cells, hinting on the differences in expression profile despite all sharing some level of expression of common Müller glia or astrocyte markers. Expression pattern of Müller glia markers (*Glul*, *Rlbp1*, *Clu*, *Apoe*) and astrocyte markers (*S100a16*, *Ndr2*, *Mlc1*, *Aqp4*) were further inspected in Fig3.3.5.1.2-3, while the expression levels are presented in Fig3.3.5.1.4 as violin plots.

The combination of all astrocyte specific markers were mostly expressed in subclusters MG/A-1 and MG/A-3, while subclusters such as MG/A-4 expressed only *S100a16*, *Ndr2*, MG/A-6 expressed only *Mlc1*, *Aqp4* but not the others. Considering Müller glia markers were prominent in all subclusters but astrocyte makers were only consistently expressed in MG/A-1 and MG/A-3, these two subclusters were categorized as the astrocyte population in our sample, while the others were defined as Müller glia. Number of Müller glia and astrocytes were mentioned in Table3.3.3.1.

As single cell RNA-seq allows us to sequence mRNA from individual cells and create individual expression profiles, ontological analysis can be applied by averaging expression profiles of cells within certain cluster

or cell type. Since we had defined the population of Müller glia and astrocytes from the previous section, differential expression gene lists between these two cell types originating from WT and KO samples were created and performed ontological analysis using IPA.

The top 50 most significantly affected Canonical Pathways in KO Müller glia and astrocytes in comparison to that of WT were shown in Table3.3.5.2.1 and Table3.3.5.2.2 respectively. Numerous neuroinflammation and senescence related pathways were also among the most significantly affected pathways. Other neuroinflammation and senescence related pathways were shown in Table3.3.5.2.3 and Table 3.3.5.2.4 respectively, with the significance ($-\log(p\text{-value})$) and degree of up/downregulation (z-score) for both Müller glia and astrocytes displayed together.

As shown from Table3.3.3.2, astrocytes contribute to about 1% of the cell population in our single cell sample. Despite having a smaller sample size in comparison to the entire retinal tissue, strong and significant activation of neuroinflammation pathways can be observed in KO astrocytes, including “Leukocyte Extravasation Signalling”, “Interferon Signalling” and “Neuroinflammation Signalling Pathway”. In contrast, Müller glia has significant but small downregulation in all these pathways except “Interferon Signalling”. This suggests astrocytes are the major cell type related to neuroinflammatory response in the retina, consolidating the importance of gene *ABCA1* and its effects in KO conditions.

| Ingenuity Canonical Pathways | -log(p-value) | z-score |
|-----------------------------------------------------|---------------|---------|
| Osteoarthritis Pathway | 10.5 | -1.257 |
| Hepatic Fibrosis / Hepatic Stellate Cell Activation | 10.2 | NaN |
| Phototransduction Pathway | 6.89 | NaN |
| Synaptogenesis Signaling Pathway | 6.42 | 0.507 |
| Hepatic Fibrosis Signaling Pathway | 6.31 | -0.632 |
| IGF-1 Signaling | 5.06 | 0 |
| Apelin Endothelial Signaling Pathway | 5.05 | -1.698 |
| Relaxin Signaling | 5 | 0 |
| IL-8 Signaling | 4.94 | 0 |
| p53 Signaling | 4.79 | 0.277 |
| Estrogen Receptor Signaling | 4.67 | -0.686 |
| Opioid Signaling Pathway | 4.65 | 1.4 |
| Thrombin Signaling | 4.64 | 0.258 |
| Cardiac Hypertrophy Signaling | 4.45 | 0.626 |
| ILK Signaling | 4.37 | -0.426 |
| White Adipose Tissue Browning Pathway | 4.36 | 0.728 |
| GABA Receptor Signaling | 4.36 | NaN |
| Signaling by Rho Family GTPases | 4.32 | 0 |
| Gα12/13 Signaling | 4.31 | 0 |
| Endocannabinoid Cancer Inhibition Pathway | 4.26 | -0.229 |
| PTEN Signaling | 3.96 | -1 |
| Mouse Embryonic Stem Cell Pluripotency | 3.94 | -0.258 |
| Acute Phase Response Signaling | 3.85 | -1 |
| Calcium Signaling | 3.83 | 1.5 |
| RhoGDI Signaling | 3.82 | 0.5 |
| CXCR4 Signaling | 3.82 | -0.258 |
| GP6 Signaling Pathway | 3.75 | 1.5 |
| Adipogenesis pathway | 3.63 | NaN |
| RhoA Signaling | 3.58 | 1.5 |
| CREB Signaling in Neurons | 3.39 | 1.069 |
| Role of NFAT in Regulation of the Immune Response | 3.35 | -0.943 |
| cAMP-mediated signaling | 3.2 | 1.528 |
| Molecular Mechanisms of Cancer | 3.2 | NaN |
| Wnt/β-catenin Signaling | 3.18 | -1.414 |
| Human Embryonic Stem Cell Pluripotency | 3.12 | NaN |
| Glioma Invasiveness Signaling | 3.12 | -0.905 |
| STAT3 Pathway | 3.12 | 1.291 |
| Actin Cytoskeleton Signaling | 3.08 | 0.447 |
| Epithelial Adherens Junction Signaling | 3 | NaN |
| NF-κB Signaling | 3 | 1.606 |
| Apelin Liver Signaling Pathway | 2.92 | 0 |
| Neuroinflammation Signaling Pathway | 2.88 | -0.229 |
| BEX2 Signaling Pathway | 2.88 | -0.302 |
| Thyroid Cancer Signaling | 2.88 | -0.905 |
| Glutamate Receptor Signaling | 2.84 | 0 |
| Phospholipase C Signaling | 2.84 | 0.471 |
| Integrin Signaling | 2.84 | 0.688 |
| Role of NFAT in Cardiac Hypertrophy | 2.82 | 0.894 |
| AMPK Signaling | 2.82 | 1.291 |
| Corticotropin Releasing Hormone Signaling | 2.79 | -0.535 |

Table3.3.5.2.1 Top 50 most significantly affected Canonical Pathways in differential expression between all Müller glia from KO vs WT samples. Yellow and green highlights label pathways related to neuroinflammation and senescence respectively.

| Ingenuity Canonical Pathways | -log(p-value) | z-score |
|-----------------------------------------------------|---------------|---------|
| Osteoarthritis Pathway | 9.18 | -0.2 |
| Calcium Signaling | 5.77 | 0.243 |
| Hepatic Fibrosis / Hepatic Stellate Cell Activation | 5.41 | NaN |
| Corticotropin Releasing Hormone Signaling | 5.3 | 0.243 |
| White Adipose Tissue Browning Pathway | 4.77 | 0 |
| CREB Signaling in Neurons | 4.69 | -0.277 |
| Synaptogenesis Signaling Pathway | 4.57 | 0.784 |
| GABA Receptor Signaling | 4.44 | NaN |
| Opioid Signaling Pathway | 4.02 | 0.447 |
| Neuroinflammation Signaling Pathway | 3.61 | 0.471 |
| Netrin Signaling | 3.29 | -0.333 |
| Phototransduction Pathway | 3.24 | NaN |
| Cardiomyocyte Differentiation via BMP Receptors | 3.23 | 0 |
| nNOS Signaling in Skeletal Muscle Cells | 3.22 | NaN |
| cAMP-mediated signaling | 3.19 | -0.728 |
| Agranulocyte Adhesion and Diapedesis | 3.19 | NaN |
| G-Protein Coupled Receptor Signaling | 3.03 | NaN |
| Relaxin Signaling | 2.96 | -0.816 |
| GP6 Signaling Pathway | 2.91 | 0 |
| Synaptic Long Term Depression | 2.85 | 1.291 |
| Virus Entry via Endocytic Pathways | 2.78 | NaN |
| Hepatic Fibrosis Signaling Pathway | 2.71 | 0.6 |
| Role of NFAT in Cardiac Hypertrophy | 2.69 | 0.728 |
| Apelin Liver Signaling Pathway | 2.68 | -1.342 |
| Melanocyte Development and Pigmentation Signaling | 2.68 | -0.707 |
| Endocannabinoid Neuronal Synapse Pathway | 2.63 | 0.302 |
| Actin Cytoskeleton Signaling | 2.61 | 1.069 |
| TR/RXR Activation | 2.49 | NaN |
| Signaling by Rho Family GTPases | 2.47 | 0.5 |
| Androgen Signaling | 2.42 | 1.414 |
| Gustation Pathway | 2.41 | NaN |
| Glutamate Receptor Signaling | 2.37 | NaN |
| ERK5 Signaling | 2.36 | 0 |
| Caveolar-mediated Endocytosis Signaling | 2.33 | NaN |
| AMPK Signaling | 2.31 | 0 |
| Type II Diabetes Mellitus Signaling | 2.27 | -0.447 |
| Gap Junction Signaling | 2.25 | NaN |
| Endocannabinoid Cancer Inhibition Pathway | 2.24 | 0.577 |
| RhoGDI Signaling | 2.24 | -1.155 |
| Neurotrophin/TRK Signaling | 2.22 | 0.378 |
| Circadian Rhythm Signaling | 2.21 | NaN |
| Cardiac Hypertrophy Signaling | 2.19 | -0.535 |
| Wnt/Ca+ pathway | 2.17 | 0.378 |
| Glycogen Degradation II | 2.09 | NaN |
| Estrogen Receptor Signaling | 2.08 | 1.091 |
| Apelin Endothelial Signaling Pathway | 2.06 | 1.667 |
| p53 Signaling | 2.05 | -0.816 |
| ILK Signaling | 2.04 | 0.832 |
| Breast Cancer Regulation by Stathmin1 | 2.04 | NaN |
| Human Embryonic Stem Cell Pluripotency | 2.01 | NaN |

Table3.3.5.2.2 Top 50 most significantly affected Canonical Pathways in differential expression between all astrocytes from KO vs WT samples. Yellow and green highlights label pathways related to neuroinflammation and senescence respectively.

| Ingenuity Canonical Pathways | MG: -log(p-value) | MG: z-score | Astro: -log(p-value) | Astro: z-score |
|---------------------------------------------------------------------------|----------------------|----------------|-------------------------|-------------------|
| CXCR4 Signaling | 3.82 | -0.258 | 1.09 | 2.121 |
| Leukocyte Extravasation Signaling | 1.55 | -2 | 1.91 | 1.941 |
| Chemokine Signaling | 1.43 | -0.707 | 0.485 | 1 |
| Interferon Signaling | 1.56 | 2.236 | 1.4 | 1 |
| T Cell Exhaustion Signaling Pathway | 1.67 | -0.302 | 0 | 1 |
| Role of NFAT in Regulation of the Immune Response | 3.35 | -0.943 | 1.19 | 0.905 |
| IL-8 Signaling | 4.94 | 0 | 1.23 | 0.577 |
| Neuroinflammation Signaling Pathway | 2.88 | -0.229 | 3.61 | 0.471 |
| Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis | 2.67 | NaN | 1.25 | NaN |
| Tight Junction Signaling | 2.16 | NaN | 1.08 | NaN |
| TNFR1 Signaling | 1.5 | 0.447 | 0.265 | NaN |
| Virus Entry via Endocytic Pathways | 1.87 | NaN | 2.78 | NaN |

Table3.3.5.2.3 Neuroinflammation related canonical pathways affected in Müller glia (MG) and astrocytes (Astro) of KO samples, sorted by descending astrocyte z-score.

| Ingenuity Canonical Pathways | MG: -log(p-value) | MG: z-score | Astro: -log(p-value) | Astro: z-score |
|------------------------------|----------------------|----------------|-------------------------|-------------------|
| HMGB1 Signaling | 2.23 | -0.775 | 0.26 | 1.633 |
| Androgen Signaling | 1.86 | 1.134 | 2.42 | 1.414 |
| Estrogen Receptor Signaling | 4.67 | -0.686 | 2.08 | 1.091 |
| mTOR Signaling | 2.56 | -0.775 | 0.46 | 1 |
| Thrombin Signaling | 4.64 | 0.258 | 1.72 | 0.632 |
| BEX2 Signaling Pathway | 2.88 | -0.302 | 0.807 | 0.447 |
| Calcium Signaling | 3.83 | 1.5 | 5.77 | 0.243 |
| AMPK Signaling | 2.82 | 1.291 | 2.31 | 0 |
| Protein Kinase A Signaling | 2.51 | 1.528 | 1.99 | 0 |
| PTEN Signaling | 3.96 | -1 | 1.07 | -0.378 |
| cAMP-mediated signaling | 3.2 | 1.528 | 3.19 | -0.728 |
| p53 Signaling | 4.79 | 0.277 | 2.05 | -0.816 |

Table3.3.5.2.4 Senescence related canonical pathways affected in Müller glia (MG) and astrocytes (Astro) of KO samples, sorted by descending astrocyte z-score.

In the case for senescence, Table 3.3.5.2.4 showed some pathways are being upregulated in astrocytes, such as “Androgen Signaling”, “Estrogen Receptor Signaling” and “Thrombin Signaling”, while others such as “cAMP-mediated signaling”, “p53 Signaling” were slightly downregulated. However, up/downregulation of senescence pathways in Müller glia is almost opposite to that of astrocytes with high significance and degree. This data shows both Müller glia and astrocytes may experience senescence in the KO condition, but the pathways and mechanisms of how they are being affected may be different. Cell type specific activation of certain pathways could provide important future leads to potential explanation of how different cell types respond differently during disease development and progression.

3.3.6 Retinal Ganglion Cells subclusters

3.3.6.1 Identifying subcluster

Retinal ganglion cells (RGCs) have a significantly smaller cell population compared to other cell types.

Thus, it was much more challenging to define a RGC cluster without over clustering other cell types in the main plot. As shown from Fig3.3.3.7E-H, RGC markers mainly express in the “RGC” labelled cluster, as well as the surrounding clusters “Unidentified-2” and “Unidentified-6”. To preserve as much RGCs as possible, cells from cluster “RGC”, “Unknown-2” and “Unknown-6” were grouped together to perform subclustering. 16 subclusters were created with a granularity of 1.6, projected as the tSNE diagram in Fig3.3.6.1.1.

To determine the RGC population, additional RGC specific markers (*Chrn6*, *Cntnap2*, *Sncg*, *Thy1*, *Resp18*, *Slc17a6*, *Stmn2*, *Chrn3*) expression patterns were shown in Fig3.3.6.1.2 and the expression level were shown in the violin plot Fig3.3.6.1.3. Data suggested high expression level of RGCs marker genes were found in subcluster 11 and 16, thus classified as separate RGC subcluster, RGC-2 and RGC-1 respectively.

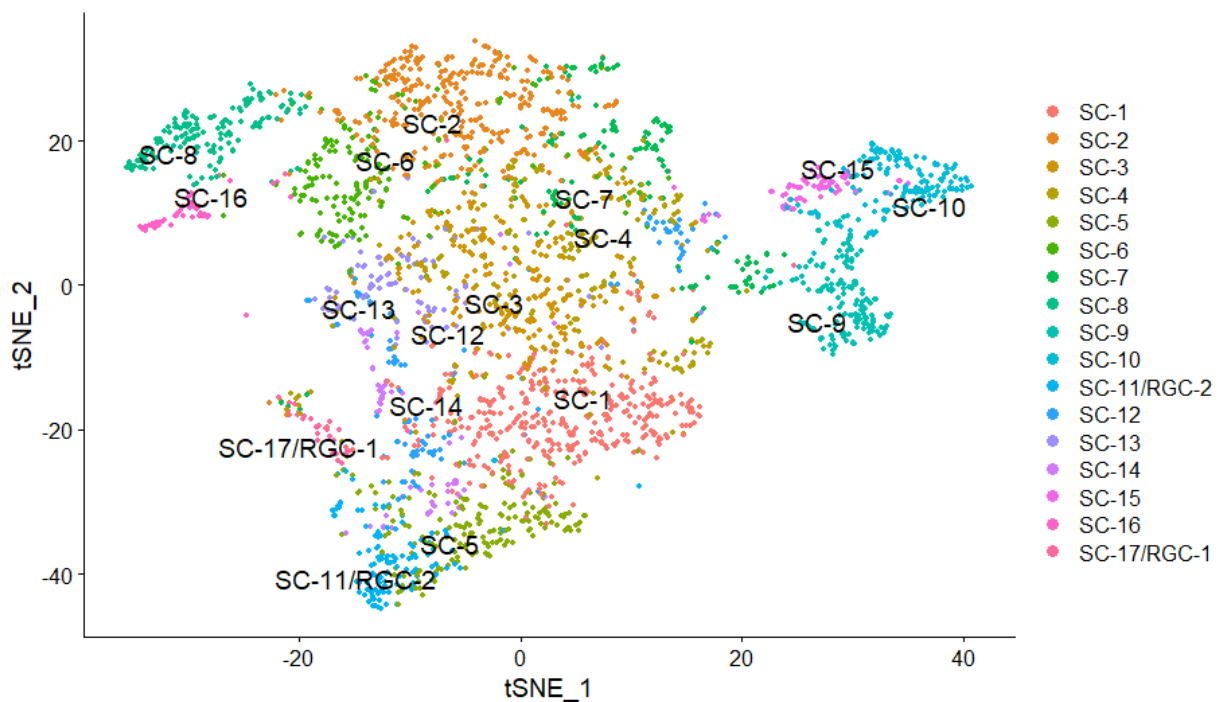


Fig3.3.6.1.1 tSNE subclustering of “RGC”, “Unidentified-2” and “Unidentified-6” from the main plot

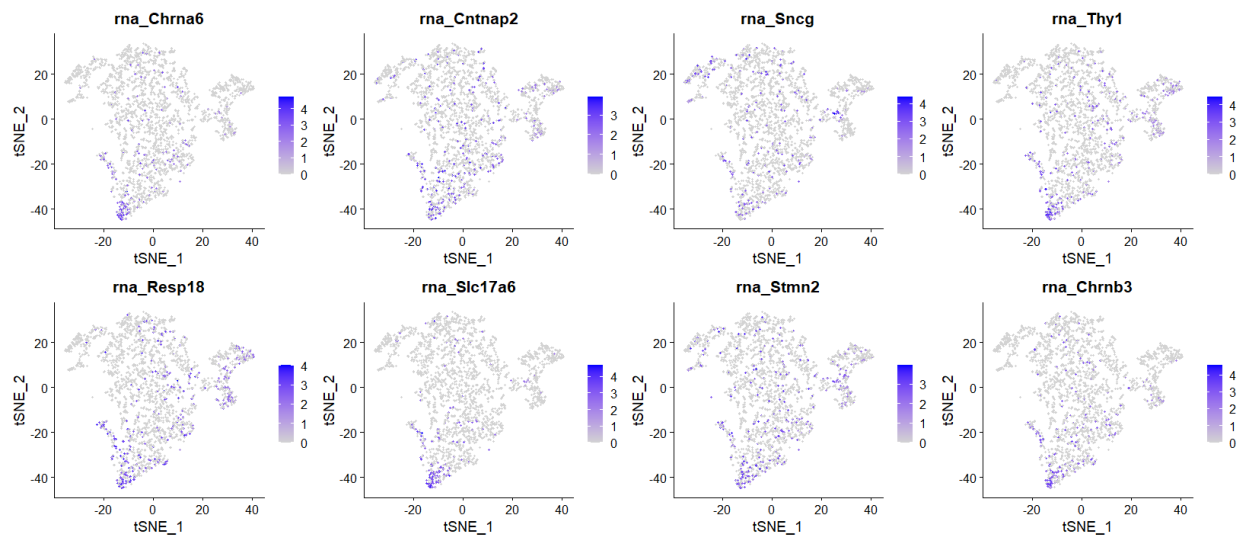


Fig3.3.6.1.2 RGC specific markers expression in tSNE subclusters

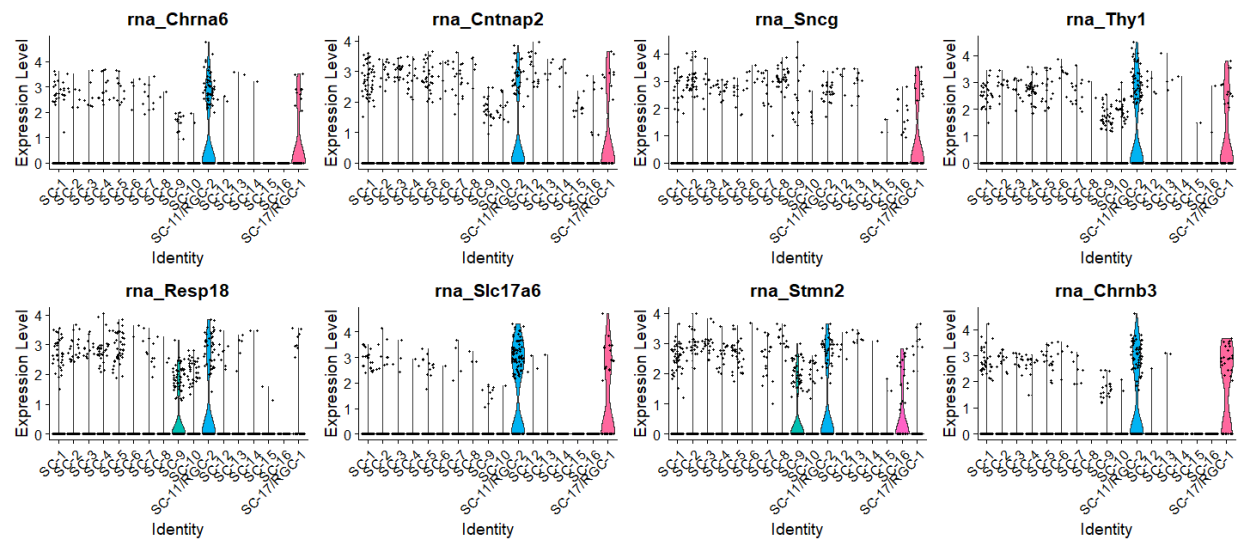


Fig3.3.6.1.3 Violin Plot of RGC specific markers expression in tSNE subclusters

3.3.6.2 RGC-subtype ontological analysis

As the clustering algorithm created 2 sub-clusters of RGCs, expressional differences between the two sub-clusters as well as difference among WT and KO cells within those clusters became of interest.

Although both clusters showed expression of typical RGC markers, differential expression between RGC-1 and RGC-2 showed some specific markers for these RGC sub-types. As shown in Fig.3.3.6.2.1a, RGC-1 exclusively expressed *Baz2b*, *Gng5*, *Plpp3*, *Rps26* while RGC-2 expressed *Cartpt*, *Scd2*, *Srinc3*, *Tmem130*. These novel markers for RGC subtype may help facilitate the identification of the respective RGC subtype within *in-vivo* samples.

Cells of both RGC-1 and RGC-2 were plotted using tSNE projection in Fig.3.3.6.2.1b, while Fig.3.3.6.2.1c-d separated these cells by their WT or KO origin respectively. As shown from these figures and Fig.3.3.6.2.1e, number of RGCs in RGC-1 was significantly reduced in KO compared to WT (* $P < 0.01$, Wilcoxon rank-sum test). This suggests a depletion of a specific RGC subtype under the influence of *ABCA1* KO.

Ontological analysis was performed by comparing KO vs WT cells in RGC-1. *Disease and Functions* affected in the KO population were shown in Table 3.3.6.2.1. Functions related to “Neoplasia”, “Degranulation”, “Cerebral disorder” and “Neurological disorder” were activated, while pathways related to “Axonogenesis”, “Quantity/ Differentiation/ Formation of retinal cells/ neurons” were significantly downregulated. This data strongly suggests that RGC-1 experienced degeneration, deterioration, and depletion phenotype that is typical in multitude of neurological disease, likely due to the effects of neuroinflammation activity and environment within KO retinae.

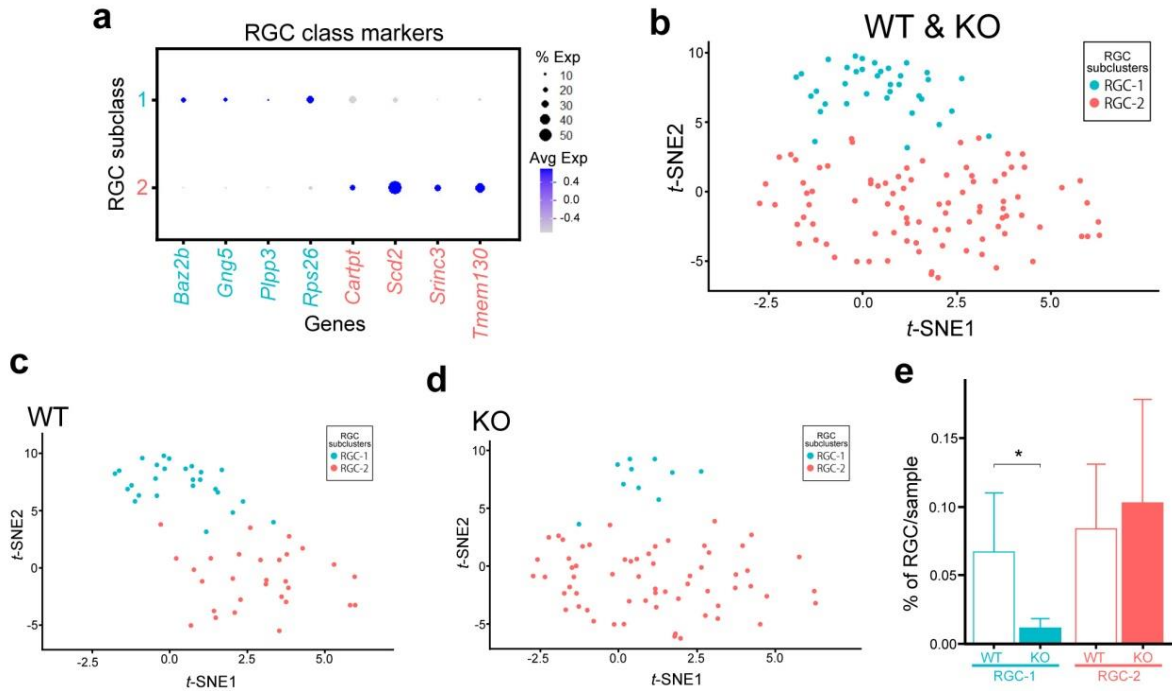


Fig3.3.6.2.1a) Dot Plot of RGC subtype markers. b) RGCs tSNE plot. c) WT only RGCs. d) KO only RGCs. e) RGC-1 and RGC-2 percentage per sample.

| Diseases or Functions Annotation | p-value | z-score |
|-----------------------------------|----------|---------|
| Neoplasia of cells | 1.07E-49 | 1.475 |
| Degranulation of cells | 8.32E-18 | 1.221 |
| Degranulation | 6.39E-18 | 1.1 |
| Cerebral disorder | 2.18E-24 | 1.056 |
| Apoptosis | 1.01E-54 | 0.798 |
| Progressive neurological disorder | 1.09E-32 | 0.363 |
| Congenital neurological disorder | 1.72E-15 | 0.126 |
| Synaptic transmission of cells | 1.2E-15 | -0.111 |
| Formation of brain | 1.11E-30 | -0.177 |
| Differentiation of nervous system | 1.65E-30 | -0.245 |
| Formation of retinal cells | 1.78E-16 | -0.476 |
| Axonogenesis | 1.4E-17 | -0.786 |
| Morphogenesis of neurons | 1.2E-31 | -0.866 |
| Neuritogenesis | 5.28E-30 | -0.866 |
| Quantity of retinal cells | 2.46E-17 | -0.902 |
| Differentiation of neurons | 1.17E-24 | -0.908 |
| Development of neurons | 3.03E-40 | -1.32 |
| Quantity of cells | 1.38E-45 | -1.455 |
| Quantity of neurons | 2.83E-28 | -1.859 |
| Quantity of nervous tissue | 4.74E-29 | -1.877 |

Table3.3.6.2.1 Disease and Functions pathway activated in RGC-1 cells of KO samples.

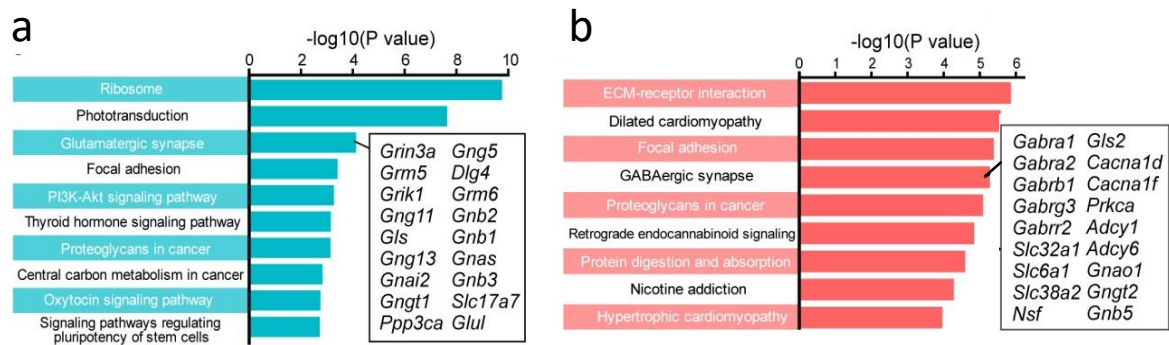


Fig3.3.6.2.2 KEGG analysis for a) enrichment and b) under-enrichment of genes expressed in RGC-1.

To further consolidate the identity of the RGC subtype experiencing neurodegeneration in the KO model, *Kyoto Encyclopedia of Genes and Genomes (KEGG)* (Kanehisa et al., 2008) was utilized by analyzing the highly expressed gene list of RGC-1 cluster. As shown in Fig3.3.6.2.2a, genes such as *Grin3a*, *Gng5* were enriched in RGC-1, which belongs to the “Glutamatergic synapse” pathway, while Fig3.3.6.2.2b shows genes such as *Gabra1*, *Gngt2* which belongs to the “GABAergic synapse” pathway being under-enriched. As genes such as *Grin3a* were downregulated in KO RGC-1 cells, this data suggests RGC-1 subtype in the retina is particularly sensitive to neuronal damage induced by the alteration or malfunction of the glutamatergic synapse, such as excitotoxicity induced neurodegeneration and apoptosis.

3.3.7 Immune cells subcluster

Ontological data analysis from previous sections suggested strong evidence of neuroinflammation in KO retinal tissues, thus piqued the interest of investigating immune response within the scRNA data.

As suggested by the main integrated data plot in Fig3.3.3.3, an “Immune cell” cluster can be found on the top of the plot, distanced away from most of other cell types. To further investigate the composition of that cluster, subclustering was performed and projected into the tSNE plot in Fig3.3.7.1A. 5 subclusters were created, with KO samples having significantly higher number of cells in all clusters in subclusters were created, with KO samples having significantly higher number of cells in all clusters in comparison to WT samples, as shown in Fig3.3.7.1B-D.

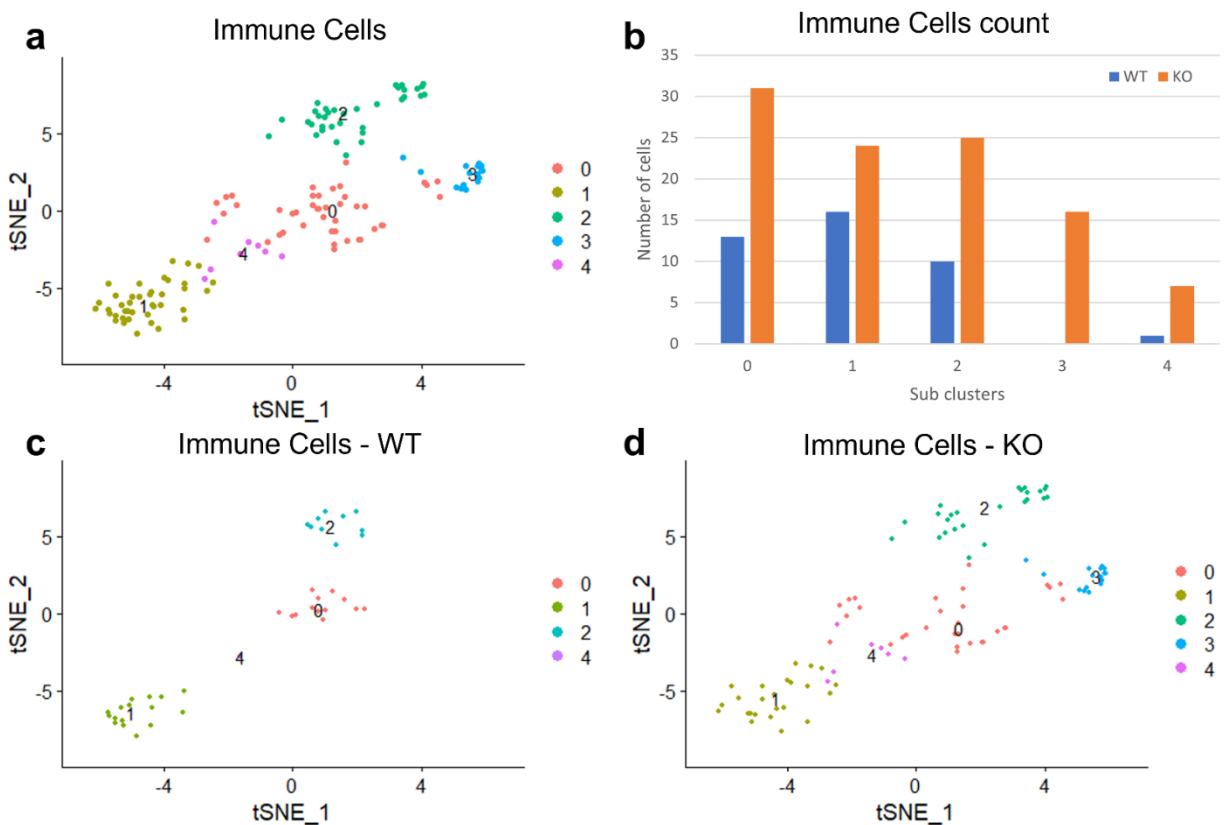


Fig3.3.7.1a) tSNE plot of Immune cell cluster. b) Immune cells cluster cell counts. c-d) tSNE plot of Immune cell cluster from c) WT and d) KO samples.

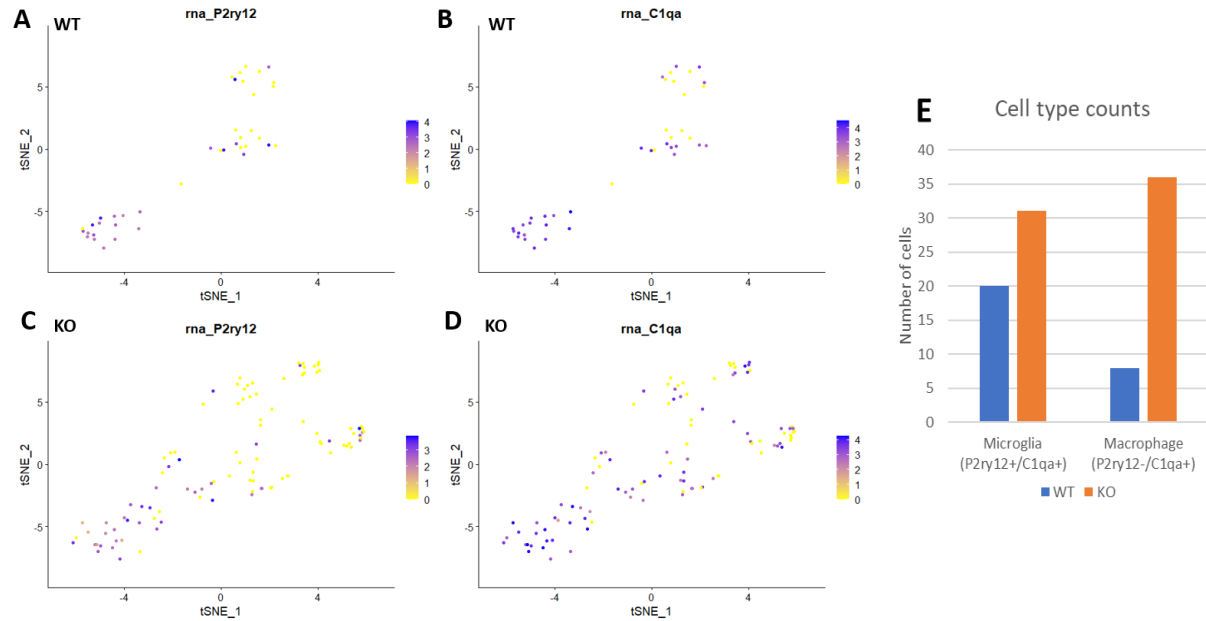


Fig3.3.7.2 Expression levels of *P2ry12* and *C1qa* in Immune Cell cluster of WT and KO samples.

Fig3.3.7.2 shows the level of expression of signature genes *P2ry12* and *C1qa* from WT and KO samples. Previous studies (Jiao et al., 2018; Masuda et al., 2020; van Hove et al., 2019) suggested *P2ry12*+/C1qa+ cells label inactive residential microglia in the retina, while loss of expression of *P2ry12* and the combinational marker of *P2ry12*-/C1qa+ suggests the activation of microglia or invasion of active macrophages. Observed from the figure, all cells from subcluster-1 expressed both *P2ry12* and *C1qa*. However, cells from the rest of the subclusters expressed high level of *C1qa* but not *P2ry12*, and majority of cells from those subclusters originated from KO samples. Altogether, evidence of macrophage invasion and microglia activation from the immune cell cluster provided strong evidence of neuroinflammation and immune response activation in *ABCA1*^{-/-} mouse retinas, potentially contributing to a subtype specific RGC apoptosis or degeneration.

3.4 Discussion

3.4.1 Choice of cell number per samples

In this experiment, we have successfully established and implemented the full pipeline of single cell RNA-seq analysis to examine WT and KO mouse retina. Tissue dissociation protocols were refined with multiple iteration before applying onto the actual samples, allowing us to achieve a sample viability of over 90% in all samples prepared. Thus, it was concerning initially when we received the early stage analysis of the sequencing data from 10X *CellRanger*, as the number of “cells” detected per sample were much lower than expected despite a satisfactory sample preparation procedure.

As suggested by the human protein atlas (Tissue-based map of the human proteome) and the mouse cell atlas study (Macosko et al., 2015b), genes expression enriched or elevated in the retinal cell population is fairly low in comparison to other tissues in the body. Within the retinal tissue however, number of genes expressed may differ depending on the cell type and their functional requirements. Fig3.4.1 shows the number of differentially expressed gene detected within different cell types detected in the mouse cell atlas study for retinal tissue (Macosko et al., 2015b). Rod photoreceptors expressed the least number of enriched genes in comparison to other cell types. However, Fig3.4.2 displaying the number of cells within each cell type showed that number of rod photoreceptors in the retina is magnitudes more than the others. This distribution was also found in our single cell dataset. Thus, the majority of the actual cells captured in the single cell samples consist of low-expressing rod photoreceptors, in which the default cell calling algorithm would have discarded assuming they were empty droplets. Although the focus of our study did not include photoreceptors, extremely stringent cell selecting would result in omitting most of the cells sequenced in the data, including other cell types which has a lower sequencing depth. Our approach of forcing top 9000 cells to be included in subsequent analysis made sure we included majority of the cells sequenced, including really low expressing cells, and filter them out once they are labelled

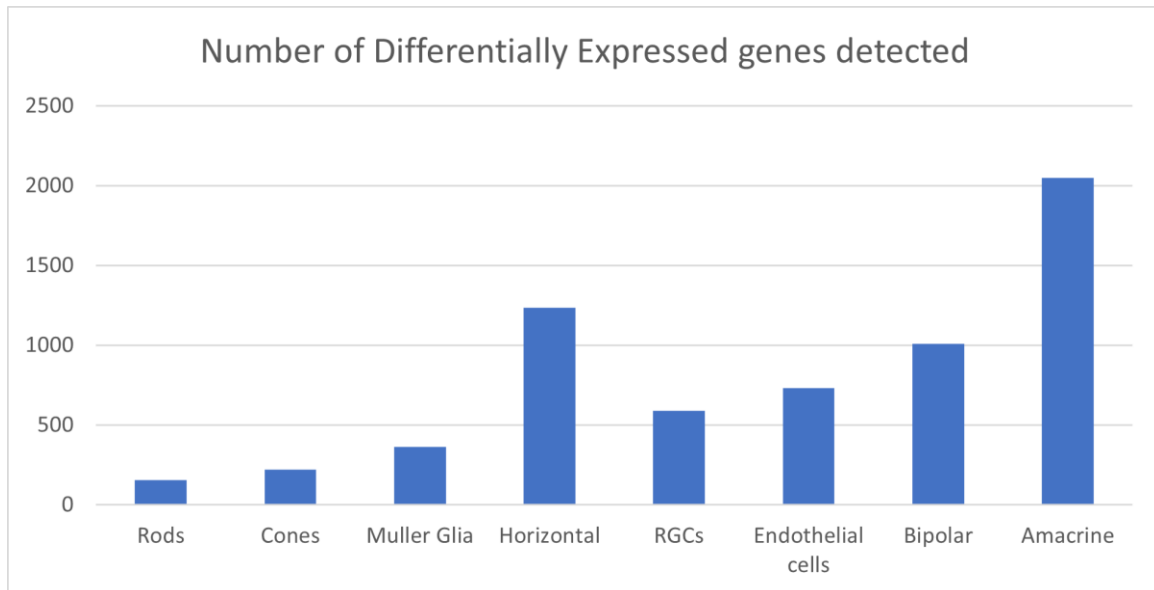


Fig3.4.1 Number of differentially expressed genes detected in different cell types (Macosko et al., 2015b).

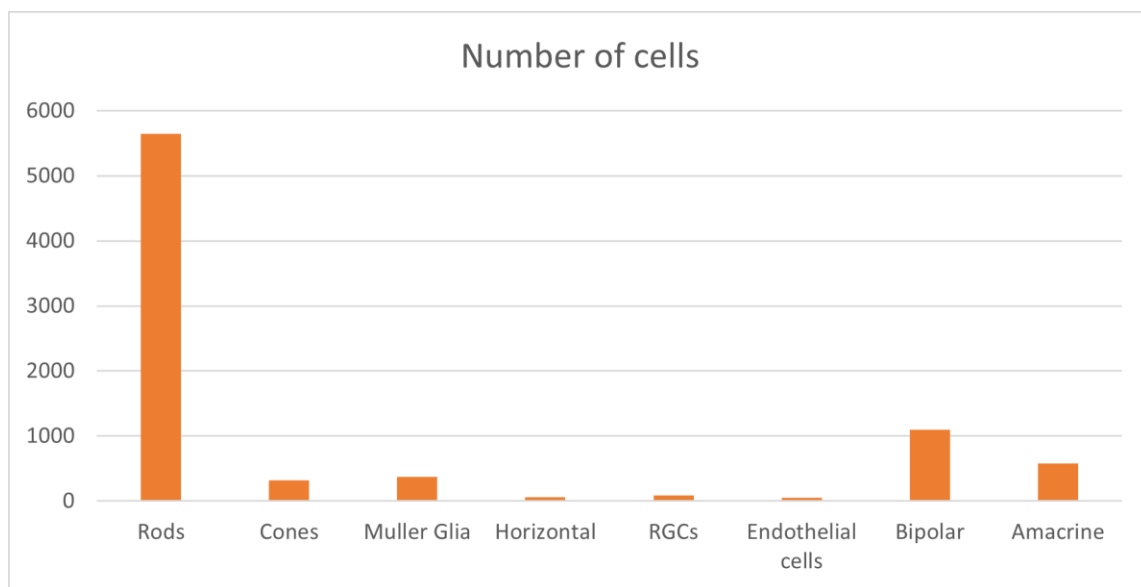


Fig3.4.2 Number of cells in different cell types (Macosko et al., 2015b).

irrelevant, such as filtering through percentage of mitochondrial gene expression and clustering them into photoreceptor clusters.

As suggested by Fig.3.2.2.1, the significant increase in number of cells included by forcing 9000 cells in comparison to the default cell calling algorithm, the number of mean reads and median genes per cell did not receive a significant drop. In fact, sample WT2, KO3 and KO4 has an increase in number of mean reads

per cell, suggesting default cell calling algorithm ignored large amount of cells which has substantial amount of reads sequenced. Subsequent analysis showed that the inclusion of more cells facilitated the clustering and identification of different clusters, including the extremely small amount of retinal ganglion cells that could have been omitted.

3.4.2 Challenges in identifying cell clusters

Rod photoreceptor dominance in the dataset not only affects the estimation of cells and the relevant statistics on average reads and genes detected, but also contribute to challenges in identifying cell types through biomarkers. Although sample viability was satisfactory, some level of background contamination might not be avoidable due to the nature of the single cell capturing technology that 10X Chromium system employs. The Drop-seq methods captures cells into droplets, but if there are RNA material contamination within the dissociated solution, mostly from the contents of dead cells, droplets may include them and be included in the expression profile of the single cell. Due to the high proportions of rod photoreceptors, majority of the contamination profile originates from this cell type. Thus expression of certain rod photoreceptor markers such as Rhodopsin (*Rho*) can be found in almost all cells at different levels. Estimation and filtering of these background contamination expression profile is hence recommended, especially in samples with higher contamination level, which will be demonstrated in the next chapter.

Gene dropout is another limitation that this single cell experiment faces which contributes to the challenges of identifying discrete cell types. 10X Genomics' technology expertise in capturing and sequencing large amounts of single cell per sample, but suffer from lower sequencing depth (number of genes detected) in comparison to other single cell sequencing methods such microfluidics or microwell technology. Gene dropout thus happens when certain genes are observed at moderate or low amount in

certain cells in a cluster, while other cells have completely no expression detected (Kharchenko et al., 2014).

This event is particularly prominent in single cell expression profiling due to the stochastic nature of mRNA expression and capturing, which amplifies by the fact that there is comparatively less mRNA expression within a single cell in comparison to a full sample in bulk RNA-seq. Since the identification of cell types rely on expression of cell type markers, inferring number of cells labelled with certain cell types can be difficult if the marker genes were not captured.

Major example includes expression of *ABCA1*, the key gene of our experiment, which supposedly is the marker for astrocytes. Although genotyping data verified the expression of said gene in WT samples and no RNA found in KO samples, extremely limited reads of *ABCA1* was detected across all samples. Thus the use of combinational markers is significantly important in identifying cell type labels for clusters, and the average expression levels of a gene in a cluster can be assumed to deduce expression level of such gene in all cells within the cluster. This assumption allows us to define the Müller glia and astrocyte subpopulation, as well as RGCs, microglia and macroglia cells.

Cell types which share large amount of expression markers also contribute to challenges in distinguishing cell types, namely Müller glia and astrocytes. These glial cells were well defined in their functions and characteristics within the retina and the nervous system, but their similarities lead to sharing of multiple common expression markers such as *Vim*, *Gfap*, *Glul* or *Slc1a3* (Geisert et al., 2009; Lukowski et al., 2019). Separation of these cell types should ideally be dictated by unique and specific cell markers that are expressed in one type but not the other. However, this is often not possible due to gene dropout phenomenon such that these genes are usually not detected at all. Similarly, progenitors at different stages of differentiation or cells sharing similar origins such as RGCs, amacrine and horizontal cells also faces similar issues, making the labelling process mentioned in Chapter 3.3.6.1 challenging. Labelling was

eventually achieved by assessing levels of expression of the combinational markers expression, which yielded satisfactory results that are comparable with other literatures relating to single cell mouse atlas. Future staining or imaging methods would be beneficial to verify validity of combinational markers used, but could also be an opportunity to utilize differential expression gene lists created from sub-clustering to identify novel cell type markers.

3.4.3 Neuroinflammation induced RGC loss in KO retina

Following the clustering analysis of the single cell data, ontological analysis using IPA was performed on multi-level of the dataset. Differentially expressed genes sets between WT and KO samples can be obtained on a whole sample level, but also by comparing specific cell types between the two conditions. Canonical pathway analysis all showed high levels of neuroinflammation occurring in the KO samples, signifying the absence of *ABCA1* expression in the astrocytes causes immune and inflammatory response in the retina, and activating different related pathways in different cell types. Evidence of neuroinflammation in the nervous systems has suggested to contribute to the development of Alzheimer's disease (Heneka et al., 2015), Parkinson's disease (Grotemeyer et al., 2022; Q. Wang et al., 2015), where the roles of astrocytes and microglia activation contributes to the inflammatory response which leads to toxic functions and neurodegeneration. As the retina is part of the central nervous system, connections between neuroinflammatory response and neurodegenerative disease can be insightful to investigate. Our data identified specific pathways that are highly activated in these relevant cell types, in which pathway such as *CXCR4 Signaling* were identified to have increased risk for progressive supranuclear palsy and Parkinson's disease (Bonham et al., 2018). Microglia activation was also detected as shown in Chapter 3.3.7, with an increasing number of macrophages found in KO samples, further consolidating the evidence of neuroinflammation and inflammatory response.

Downstream effect of this mechanism and the effect of the KO can be seen in the changes in number and pathways activated in the RGC population. Although the number of RGCs identified were limited, which is an accurate representation of the population distribution of this cell type, neurodegenerative pathways and related mechanisms were shown to be activated. Depletion in number along with evidence of neurodegeneration found in RGC, aligns with the findings of our collaborators studying this *ABCA1*^{-/-} model while additionally providing insights onto cell type specific gene and pathways variations and changes. In particular, the identification of injury susceptible RGC subtype invites further investigation on relation between astrocyte dysfunction and excitotoxicity related neurodegeneration.

3.5 Conclusion

In this chapter, we have implemented and performed the single cell RNA-seq analysis experiment on 3 WT and 5 *ABCA1*^{-/-} mouse retinal samples. Successful sample preparation led to satisfactory sequencing data, and the full *Seurat Integration* pipeline was implemented to conduct further bioinformatics analysis of the sequencing data. Different cell types of the retina were identified, with focuses on Müller glia, astrocytes, retinal ganglion cells (RGCs) and immune cells such as microglia and macrophages. Ontological analysis suggested activation of neuroinflammation and senescence related pathways in most cell types, and further evidence of neuroinflammation was found by the microglia activation and macrophage invasion in KO samples. Neurodegeneration pathways were activated in KO RGCs, with a particular subtype found with less in number and determined to be more vulnerable to excitotoxicity.

Chapter 4: P2Y6^{-/-} Experiment

4.1 Preface

Different types of glaucoma affect different groups of patients around the world, with majority of them suffering from Primary Open Angle Glaucoma (POAG). The most prominent feature of this disease involves the increase in Intraocular Pressure (IOP), which is the fluidic pressure built up within the eye by balancing between aqueous humor production and outflow. The mechanistic relationship between IOP increase and retinal damage is still fairly uncertain, thus the animal model supplied by our Japanese collaborators, the *P2Y6* knockout model (KO), provided an opportunity to study transcriptomic changes in retinal tissues affected by substantial increase in IOP.

In this chapter, we attempted to use single cell RNA-seq to analyse retinal tissue samples of both wild-type (WT) and *P2Y6*^{-/-} (KO) mouse. Low viability sample preparation had created highly contaminated sequencing data samples, in which we tried to implement different bioinformatics algorithm to estimate and remove background contamination expression profile. The data was then proceeded with the rest of the single cell data analysis pipeline to filter, normalize, integrate and visualize the dataset, and identified major important cell type clusters such as RGCs and Glial cells.

4.2 Tissue Preparation

Three WT and three *P2Y6*^{-/-} (KO) mouse were used for this experiment. Due to limitations during the Covid-19 pandemic, KO model mouse were not able to be transported to the UK. Hence the sample preparation and 10X Single cell experiment were performed by our collaborator in Japan with accordance to the methods mentioned in Chapter 2.2.3.2. Dissociation protocol employed was different from that was used in the *ABCA1*^{-/-} experiment, which yielded sample viability of about 30%-50% among different samples. Fig4.2 showed some example images from the Trypan blue assay, suggesting the presence of dead and clumpy cells sound in the dissociated samples. These samples were ultimately still proceeded

for scRNA experiment using the 10X Chromium Single-Cell 3'RNA-seq v3.1 kit and sequenced on the Novaseq S4 platform.

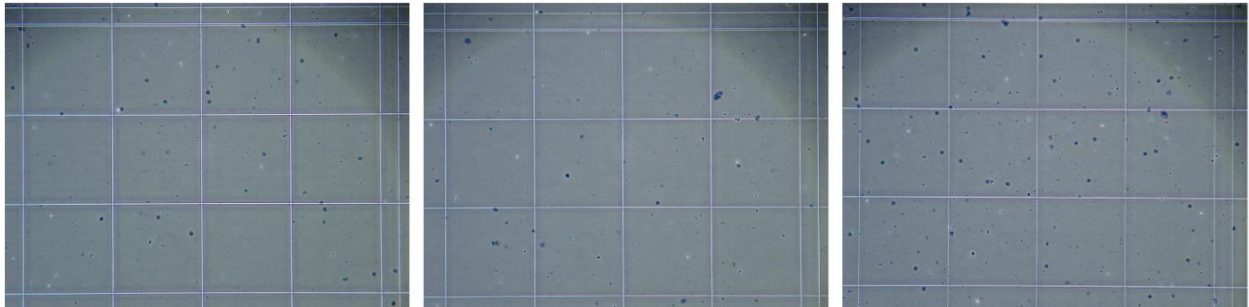


Fig4.2 Example images from Trypan blue assays of dissociated samples for P2Y6^{-/-} experiment.

4.3 Data Analysis

4.3.1 10X Genomics Cell Ranger output

Cell Ranger mkfastq output for all samples were obtained from our collaborators. Table4.3.1 summarizes the statistics for each sample, with over 95% of reads from all samples were mapped to the mouse genome, suggesting a high quality of sequencing. Other information, however, showed interesting differences in comparison with *ABCA1*^{-/-} experiment.

Fig4.3.1a-c displays the data for estimated number of cells, mean reads per cell and median number of genes per cell respectively for all samples. As the scRNA experiments were conducted at 3 different times, each with a pair of WT and KO samples, their statistics showed pairwise trends as well with high fluctuation between paired samples. The “estimated number of cells” and “mean reads/cell” showed inverse relation, suggesting the high number of “cells” are likely due to the inclusion of more empty droplets or droplets containing very little RNA content, such as that of background contamination, by the default cell-calling algorithm employed by *Cell Ranger*. Thus the actual number of single cells per sample could be theorized to be the average among all samples (~30,000). This number is still significantly higher than that of *ABCA1*^{-/-}. With consideration of a generally lower viability of samples for the *P2Y6*^{-/-}

experiment, there is a higher likelihood that these samples contain high levels of background contamination of RNA molecules from dead or dying cells. Thus a stronger and stricter quality control and filtering is required to avoid inclusion of low quality, uninformative and contaminated droplets or cells.

| | Estimated no. of cells | Mean reads/cell | Median gene/cell | Sequencing saturation (%) | % Reads mapped to genome | total genes detected |
|-----|------------------------|-----------------|------------------|---------------------------|--------------------------|----------------------|
| WT1 | 25989 | 17494 | 256 | 90.2 | 96 | 18301 |
| WT2 | 14276 | 34733 | 481 | 87.3 | 96.1 | 18744 |
| WT3 | 48627 | 10623 | 229 | 90.2 | 96 | 18971 |
| KO1 | 25495 | 20454 | 261 | 90.5 | 95.7 | 18600 |
| KO2 | 11443 | 45636 | 345 | 91.5 | 95.8 | 17683 |
| KO3 | 49480 | 10022 | 259 | 87.6 | 96.1 | 19181 |

Table4.3.1 CellRanger sequencing output summary of P2Y6^{-/-} experiment samples

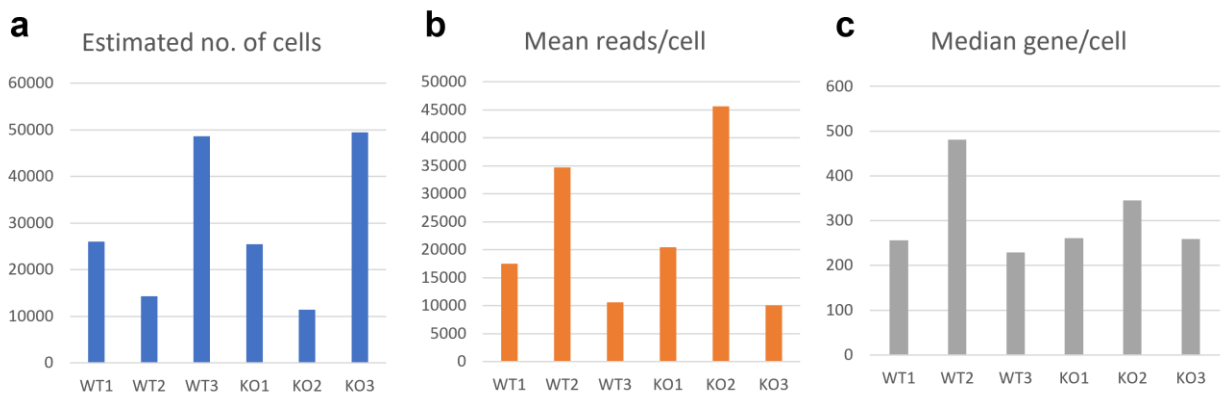


Fig4.3.1a) Estimated number of cells, b) mean reads per cell and c) median number of genes per cell in P2Y6^{-/-} samples.

4.3.2 SoupX – Contamination estimation and removal

Estimation of highly contaminated expression profile of single cells droplets can be challenging as it is indistinguishable whether RNA materials originated from a single cell, or if it was included in the droplet as RNA molecules that were suspending in the dissociated solution. SoupX (Young & Behjati, 2018) is an algorithm developed to estimate and remove the transcriptomic profile of background contamination, and was applied onto P2Y6^{-/-} samples as prefiltering quality control.

4.3.2.1 Unfiltered samples

Initially, individual samples of the *P2Y6*^{-/-} experiment were analysed by processing through the *Seurat* pipeline. *CellRanger count* function creates the expression matrix of filtered cells, which contains rows of genes identified, and columns of barcodes belonging to droplets that the cell-calling algorithm identifies as “cell”. This matrix is then processed by normalization, top 2000 most variable genes extraction, scaling, PCA, clustering using the first 15 PCs and a granularity of 1.5, and ended with a UMAP projection. Fig4.3.2.1.1-2 column A shows the UMAP projection for all 6 samples. To assess the level of contamination, expression level of rod photoreceptor marker *Rho* was inspected. As mentioned before, they are the cell type that are most abundant in the retina, thus majority of the RNA molecules that constitute the background contamination profile should originate from rod photoreceptors.

Expression levels of *Rho* in different samples were suggested to be high across all cells in all samples, as shown in Fig4.3.2.1.1-2 column B. Alongside with other rod photoreceptor markers, this consolidates the suspicion on high background contamination among all samples, as these markers should only exclusively express in rod photoreceptors and none in other cell types. SoupX *plotMarkerMap* function creates the plots shown in Fig4.3.2.1.1-2 column C, which indicate the ratio between the level of expression of *Rho* in each cell to the average value across the sample. Cells labelled in green suggests high likelihood that *Rho* expression belongs to the single cell captured in the droplet, i.e. Rod photoreceptors, which corresponds to clusters with high expression of *Rho* in Fig4.3.2.1.1-2 column B. Expression of *Rho* in other cells are thus likely due to background contamination, in which SoupX was then implemented to estimate its level and attempt to remove accordingly.

4.3.2.2 Contamination profile and fraction estimation

CellRanger count function employs their cell-calling algorithm to produce the “filtered expression matrix” of droplets regarded as cells, but also creates a “raw expression matrix” which includes all droplets that was sequenced during the 10X experiment. SoupX utilizes both matrices to estimate the expression profile

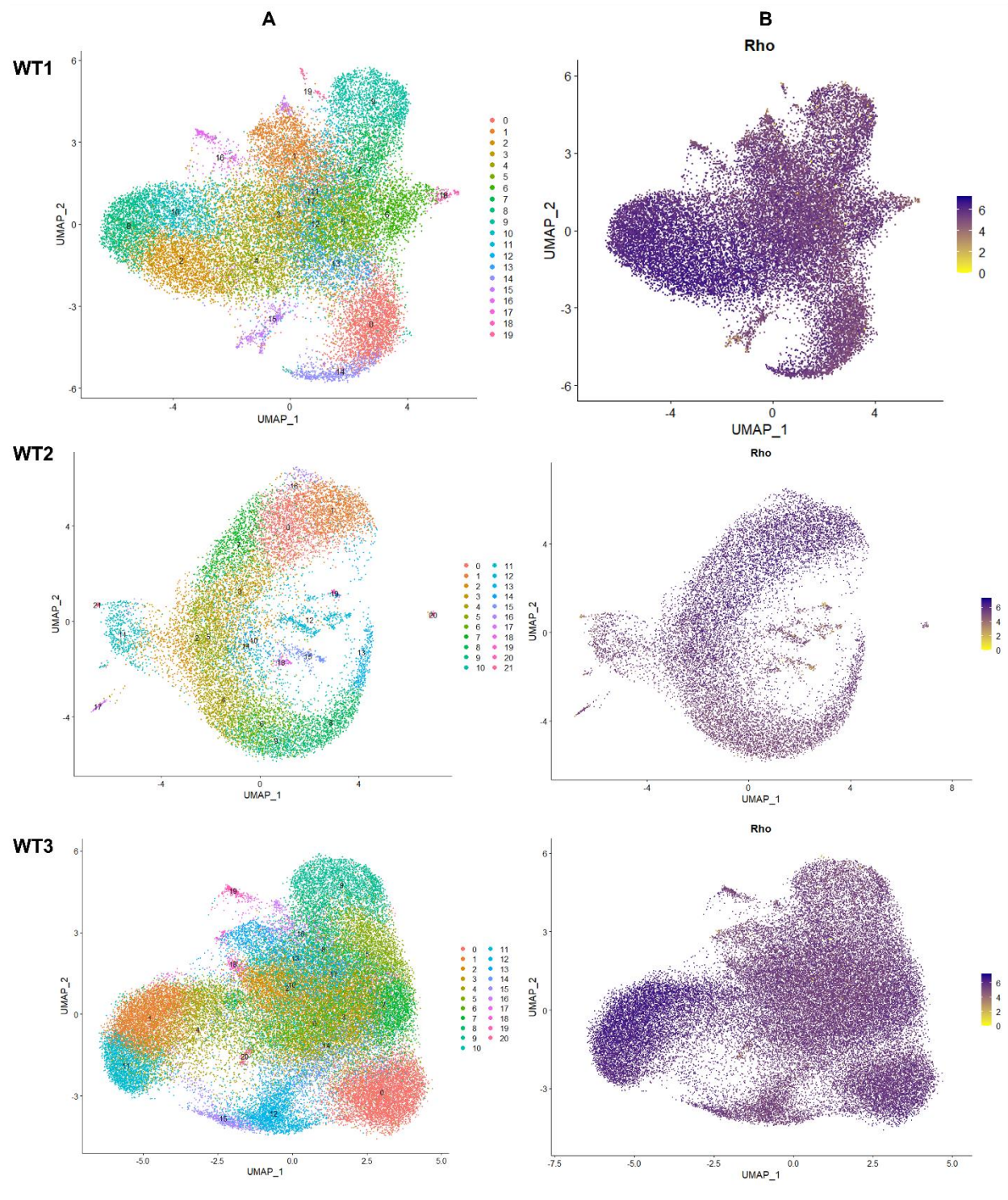


Fig4.3.2.1.1 Rows: WT samples. Column A: Original unfiltered UMAP projection. Column B: *Rho* expression.

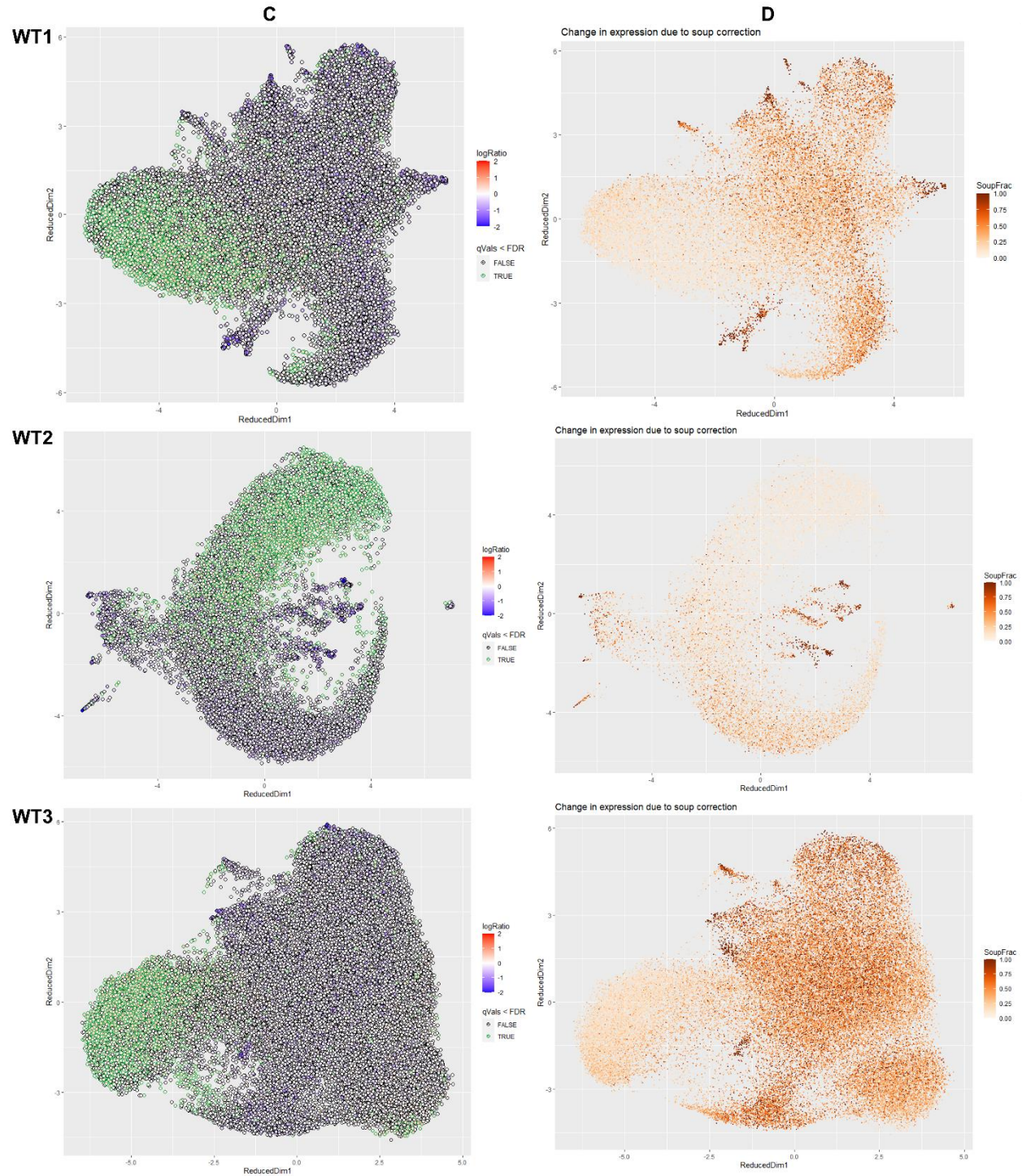


Fig4.3.2.1.1(cont) Rows: WT samples. Column C: *Rho* expression measured by the ratio between individual expression and average *Rho* expression level. Green labels signify the likelihood that *Rho* expression in that droplet belong to the cell. Column D: *Rho* expression changes after *SoupX* correction.

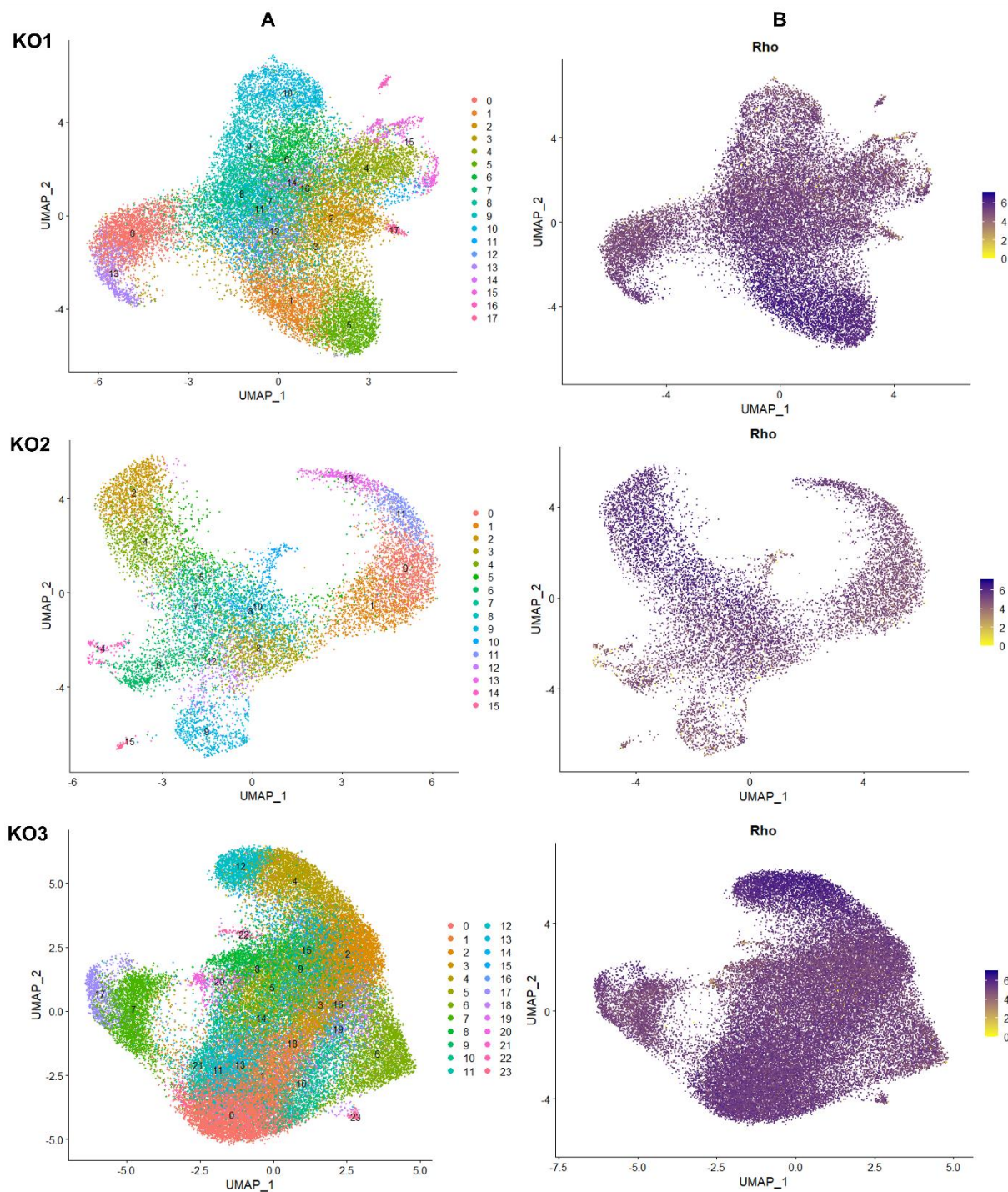


Fig4.3.2.1.2 Rows: KO samples. Column A: Original unfiltered UMAP projection. Column B: *Rho* expression.

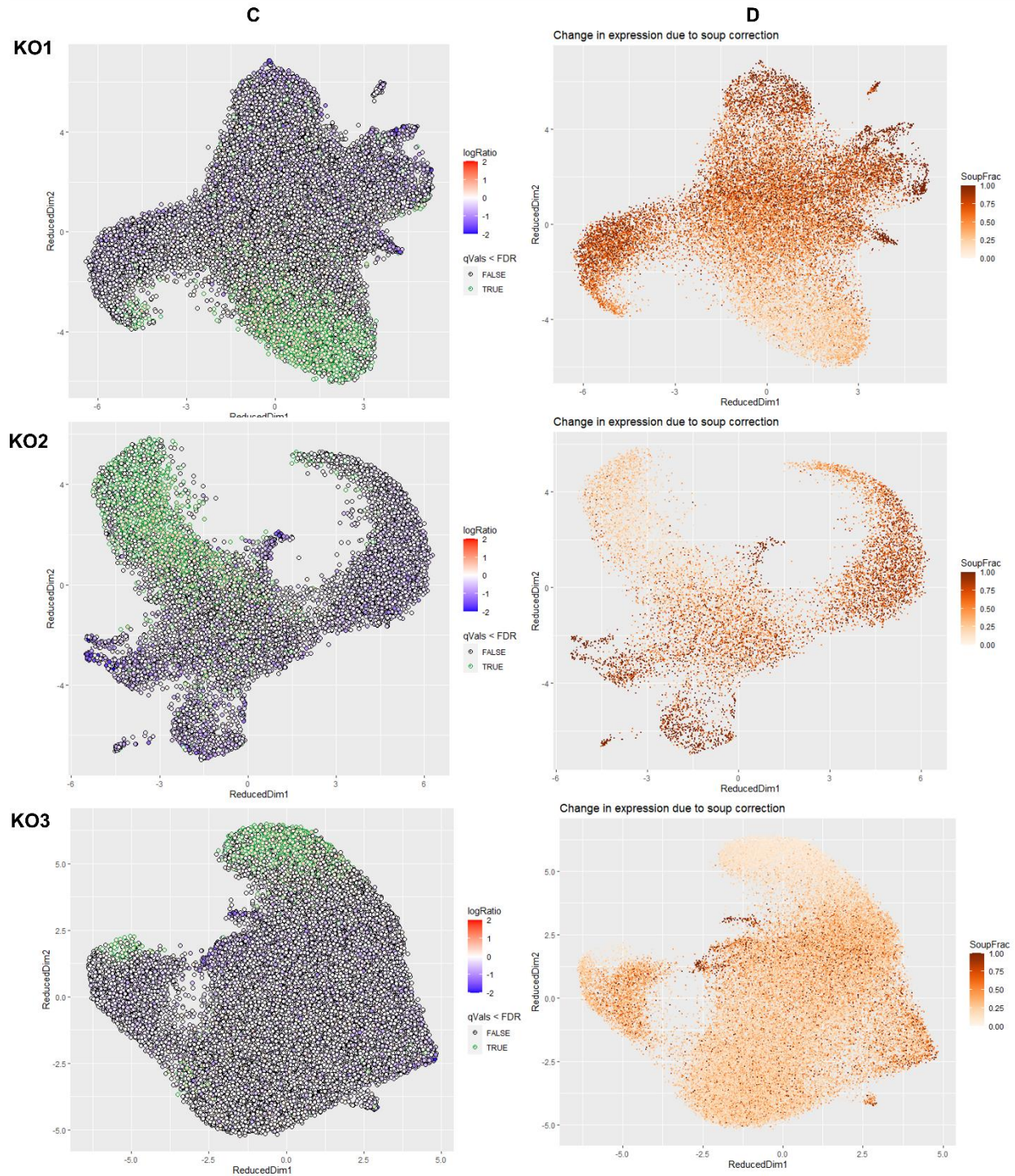


Fig4.3.2.1.2(cont) Rows: KO samples. Column C: *Rho* expression measured by the ratio between individual expression and average *Rho* expression level. Green labels signify the likelihood that *Rho* expression in that droplet belong to the cell. Column D: *Rho* expression changes after *SoupX* correction.

of the empty cells, i.e. droplets in the “raw expression matrix” that was not in the “filtered expression matrix”. *autoEstCont* function then estimates the level of contamination (ρ , p) and predicts the amount required to remove the background contamination profile. Fig4.3.2.1.2 shows the estimation of p in all 6 samples, where the solid line represents the likelihood of p being accurate, with the most probable value labelled at the red line. Besides from WT1, all other samples have a p value higher than 0.6, suggesting over 60% of expression could be contamination. This could be due to 1) an extremely high level of background contamination observed in all cells among all samples, or 2) the background expression profile is similar to the expression profile of the majority of cell types, which are the rod photoreceptors. To effectively utilize the estimation algorithm of SoupX but also preserve as much biological information hidden within the data without over-filtering, a lower peak value (indicated by the blue line in Fig4.3.2.1.2) was chosen for the samples as an estimation of the contamination fraction p .

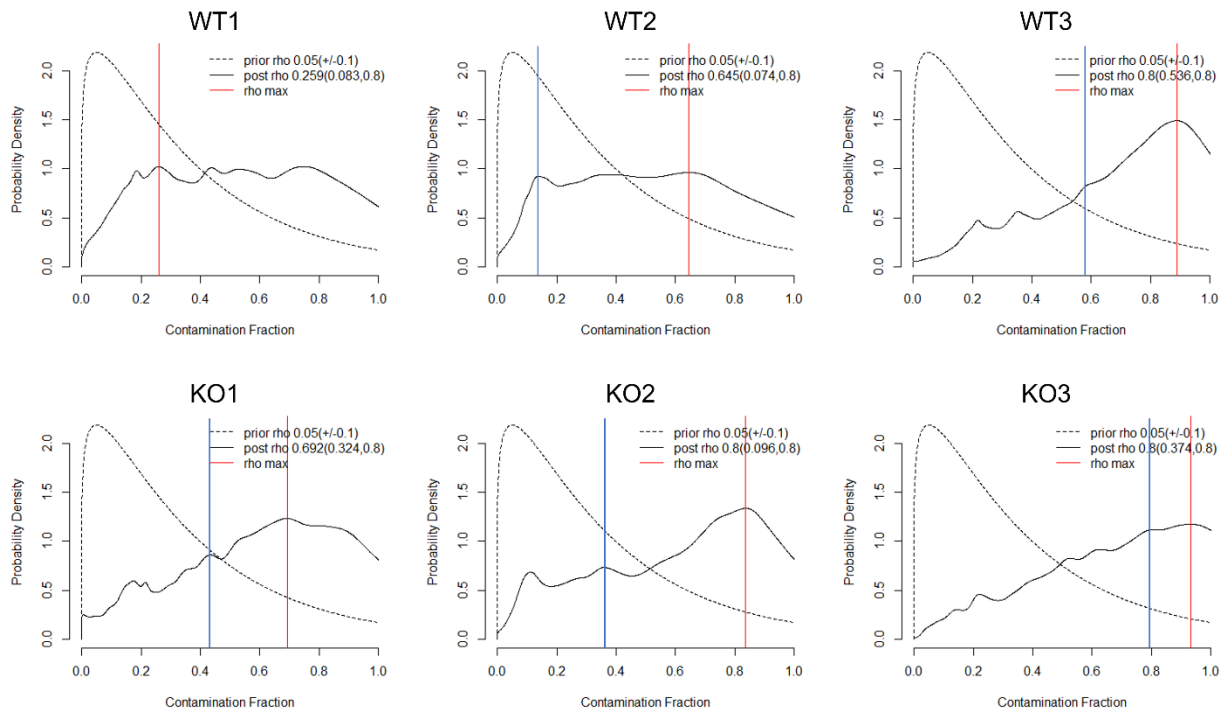


Fig4.3.2.2 Estimation plot produced by *autoEstCont* function to predict likelihood of contamination fraction being a true representation of the sample data. Red line represents maximum likelihood chosen by the algorithm, while the blue line represents the conservative value chosen.

4.3.2.3 Contamination removal

After determining the contamination factor for each sample, *adjustCounts* function was applied to remove contamination expression from the data. Fig4.3.2.1.1-2 column D showed the changes in expression of *Rho* in all 6 samples after the SoupX correction, with projection coordinates from the previous UMAP. Lighter color signifies less to no correction, while the darker color indicates a high level of contamination removal. This procedure appropriately conserved expression at regions that would be green in plots from column C, which were established to be likely rod photoreceptors, while the other cells received correction and contamination removal. Fig4.3.2.3.1-2 left column shows the new UMAP projection of the corrected expression matrix after processing through the *Seurat* analysis pipeline again. The right column showed

expression of *Rho* from the corrected data, and there were clearly more cells and clusters with lower to no expression in comparison to pre-corrected data (Fig4.3.2.1.1-2 column B). Clustering of the post-correction data also showed more distinctive and well-separated cell clusters with less overlap in comparison to pre-correction data, suggesting the SoupX process helped remove common noises and background contamination which can facilitate downstream analysis in identifying individual cell types.

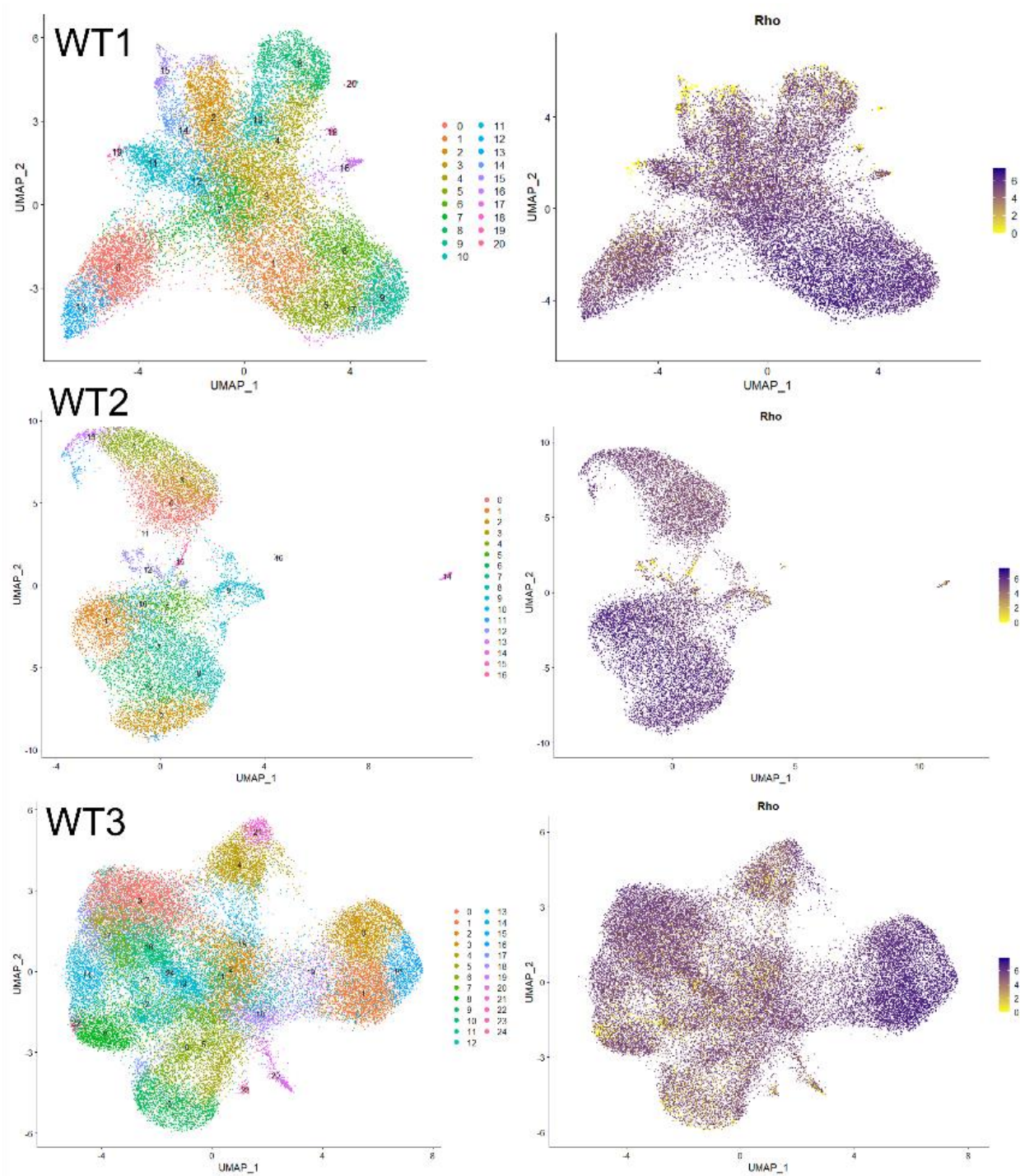


Fig4.3.2.3.1 Rows: WT samples. Left column: UMAP projection of SoupX-corrected expression data after re-analysed through the *Seurat* pipeline. Right column: *Rho* expression of corrected data in the new UMAP.

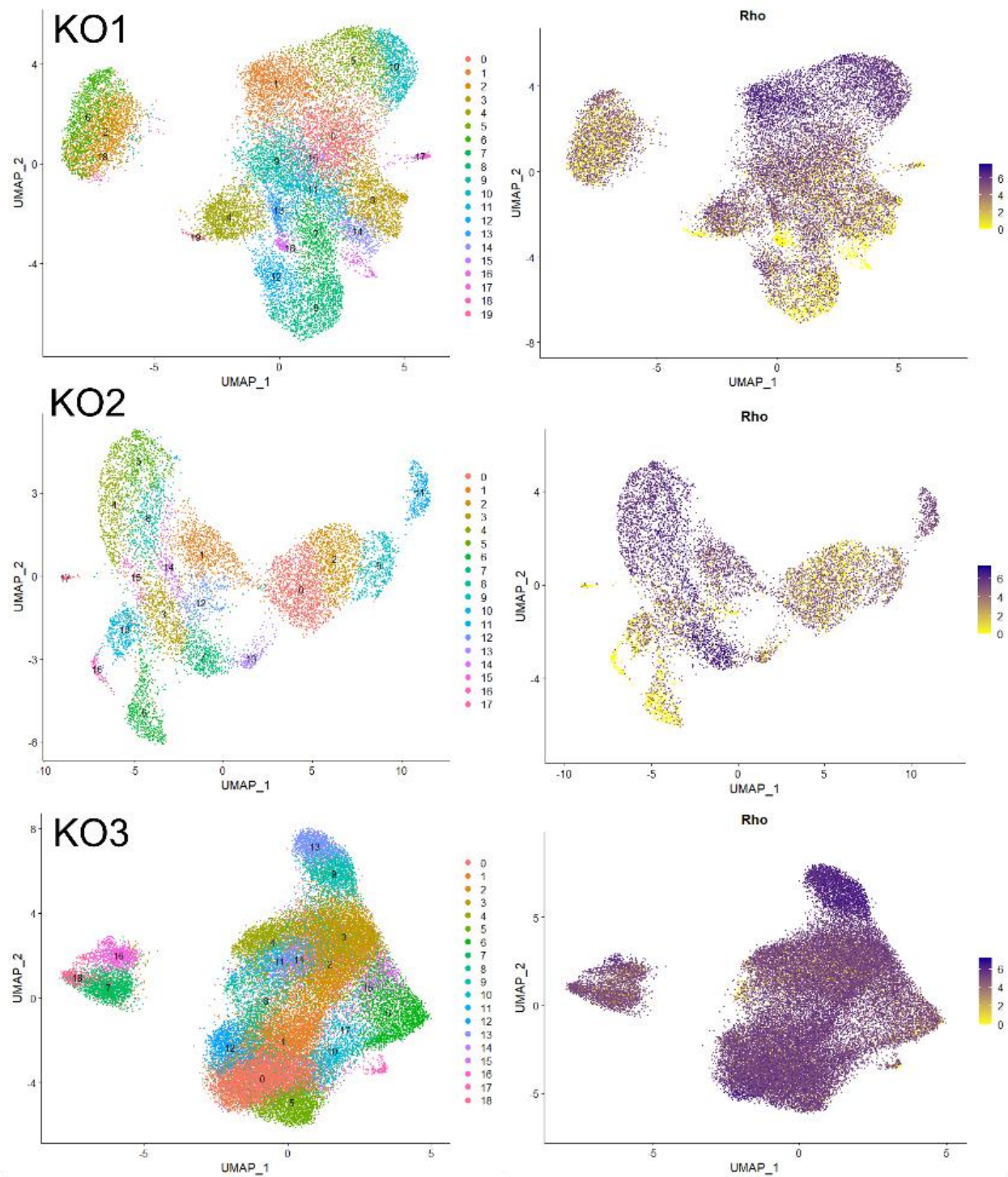


Fig4.3.2.3.2 Rows: KO samples. Left column: UMAP projection of SoupX-corrected expression data after re-analysed through the *Seurat* pipeline. Right column: *Rho* expression of corrected data in the new UMAP.

4.3.3 Seurat integration

4.3.3.1 Pre-filtering

SoupX-corrected data helped removed gene expression that originates from background contamination, and improved clustering ability of the *Seurat* algorithm. However, no cells were removed in that process, and these data points remained in the sample despite likely not containing any useful information after the SoupX process. To further refine the population included in these 6 samples, their number of genes detected, number of read counts included and percentage of mitochondrial RNA (%MT) were used as parameters for further filtering. Table4.3.3.1 showed the values chosen for minimum and maximum number of genes detected per cell, and maximum number of reads per cell to remove empty and low content cells as well as potential doublets, while maximum %MT limits the inclusion of dead or dying cells. A more stringent filter was applied to samples WT3 and KO3 by removing more low-expressing cells, such that the average number of genes and reads per cell as well as the total number of cells within these samples are more comparable to the others. The resultant average of 17,000 cells per sample were used in subsequent *Seurat* analysis.

| | Min No. of genes | Max No. of genes | Max No. of reads | Max %MT | Initial Cell No. | Filtered Cell No. | % Cells removed |
|-----|------------------|------------------|------------------|---------|------------------|-------------------|-----------------|
| WT1 | 100 | 2000 | 3000 | 20 | 25989 | 20955 | 19.37 |
| WT2 | 100 | 2000 | 3000 | 20 | 14276 | 12595 | 11.78 |
| WT3 | 200 | 1000 | 2000 | 20 | 48627 | 22681 | 53.36 |
| KO1 | 100 | 1000 | 2000 | 20 | 25495 | 18132 | 28.88 |
| KO2 | 100 | 2000 | 2200 | 20 | 11443 | 8859 | 22.58 |
| KO3 | 140 | 1000 | 1000 | 20 | 49480 | 22550 | 54.43 |

Table4.3.3.1 Filtering values used on SoupX-corrected samples to further remove low quality cells.

4.3.3.2 All samples Integration

Filtered and corrected expression profile matrices from the 3 WT and 3 P2Y6KO samples were used for the *Seurat* Integration pipeline, as mentioned in Chapter 2.3.2.5. The total of 105,772 cells were processed through the pipeline of normalization, integration, scaling, PCA, then using the top 25 PCs (Fig4.3.3.2.2) for clustering and UMAP projection, creating the plot shown in Fig4.3.3.2.3. With reference to suggestions

from the developers of *Seurat* (Tips for integrating, RPCA), reciprocal PCA (RPCA) were used instead of the default algorithm canonical correlation analysis (CCA) for integrating larger datasets with more cells. Both methods created similar results for datasets that originate from the same 10X platforms and have majority of cell types preserved across samples. Computational power and elapsed time for the *FindIntegrationAnchors* function was significantly reduced with RPCA. As suggested by Fig4.3.3.2.1, samples from both WT and KO were projected onto overlapping regions, showing effective batch effect removal and integration from the process.

Following repeated testing and evaluation, granularity of 1.1 as a resolution was used for the *FindCluster* function to create 30 different clusters, as shown in Fig4.3.3.2.3. Differential gene expression of each cluster was computed, and labelled with their respective cell identity based on high expression of retinal cell type markers. Fig4.3.3.2.4 displays the cell cluster expression of different markers by a dot plot, while Table4.3.3.2.1-2 tabulates cell types number and distribution across different samples respectively. Fig4.3.3.2.5-7 displays the expression level of different cell type markers in the integrated dataset. Majority of the cell types were dominated by rod photoreceptors as expected, with other cell types having similar relative similar proportions to regular retina and that of *ABCA1*^{-/-} experiment. Markers belonging to different cell types such as amacrine cells, horizontal cells, progenitor cells or cells undergoing cell divisions were found in their relative clusters labelled in Fig4.3.3.2.3, which were grouped under the “other” cell type label in Table 4.3.3.2.1-2 as they are not the focus of our study.

Refined subclustering of these cell types can be performed in the future should these cell types become relevant to investigating mechanism and expressions related to glaucoma development. Clusters labelled as Müller glia and RGCs have a higher percentage in comparison to the *ABCA1*^{-/-} experiment, due to the region in which their markers were highly expressed are bigger and contained more cells. Further subclustering and refinement of these specific cell types were performed in the following section to identify these cell types more accurately.

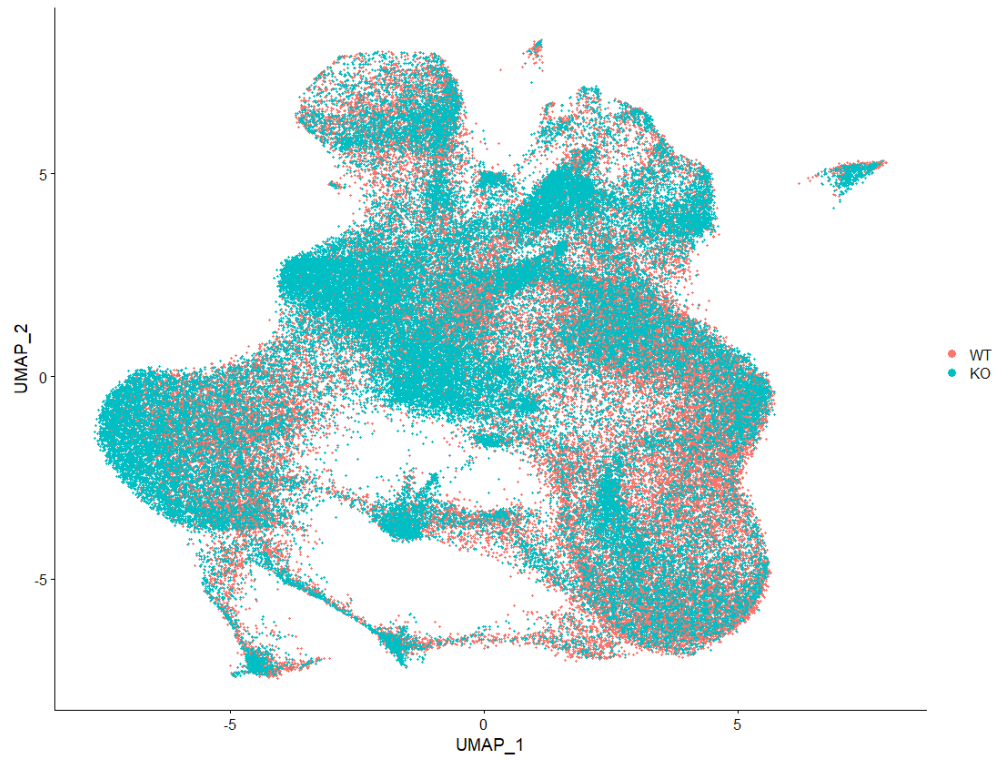


Fig 4.3.3.2.1 UMAP plot of integrated cells from 3 WT and 3 P2Y6^{-/-} dataset

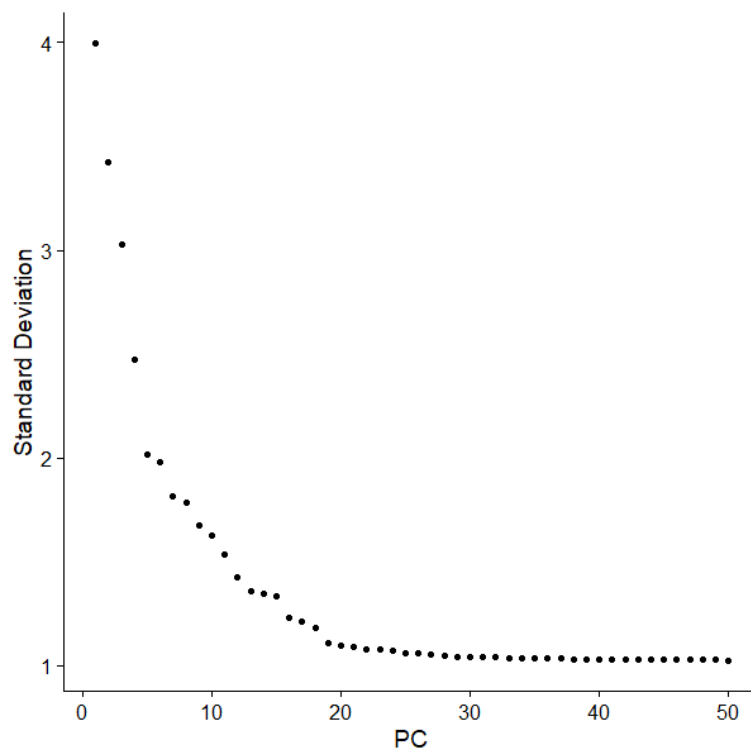


Fig4.3.3.2.2 Elbow Plot from PCA of the integrated dataset of P2Y6^{-/-} experiment.

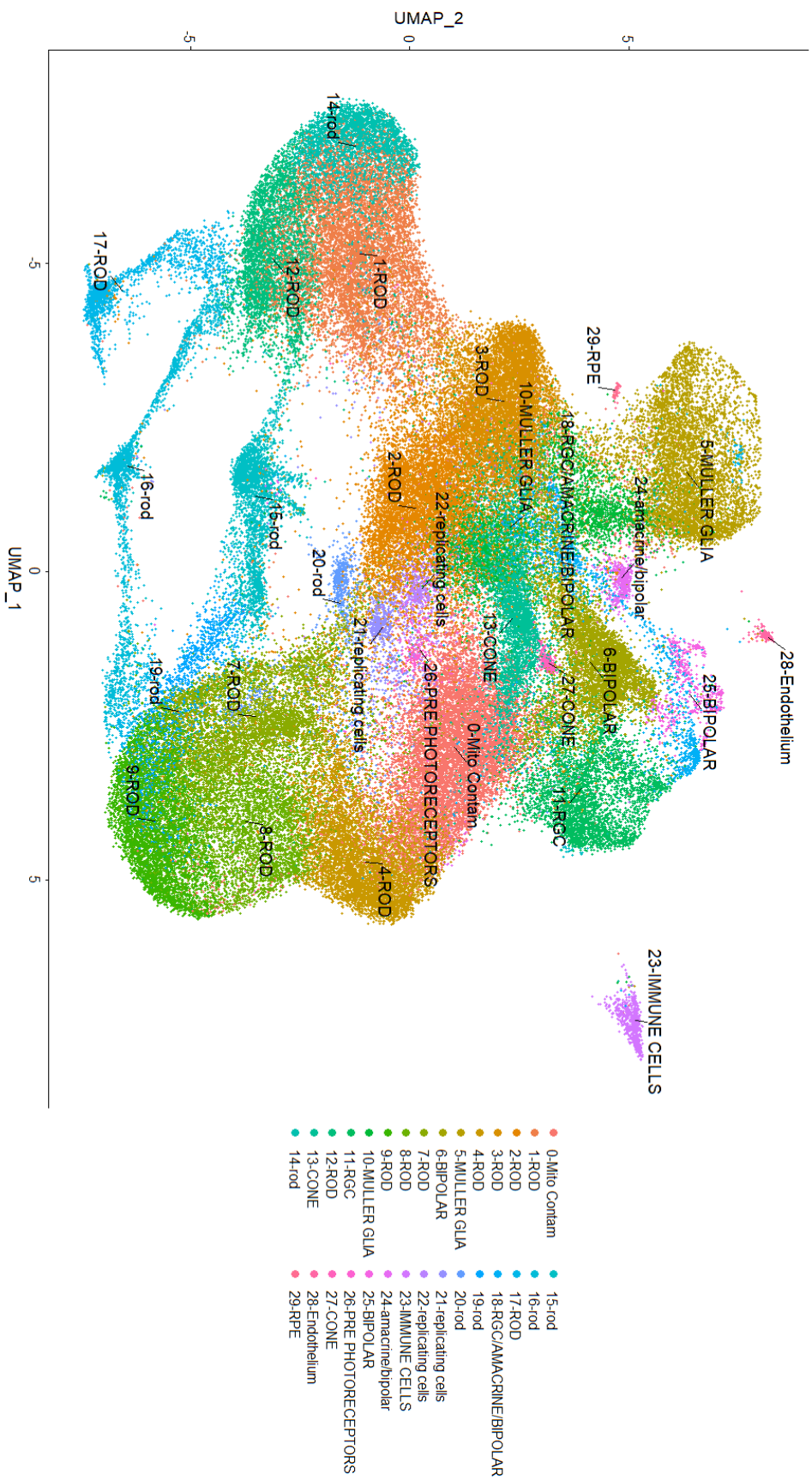


Fig4.3.3.2.3 UMAP plot of integrated cells from 3 WT and 3 P2Y6KO dataset, clustered and labelled with retinal cell type identity.

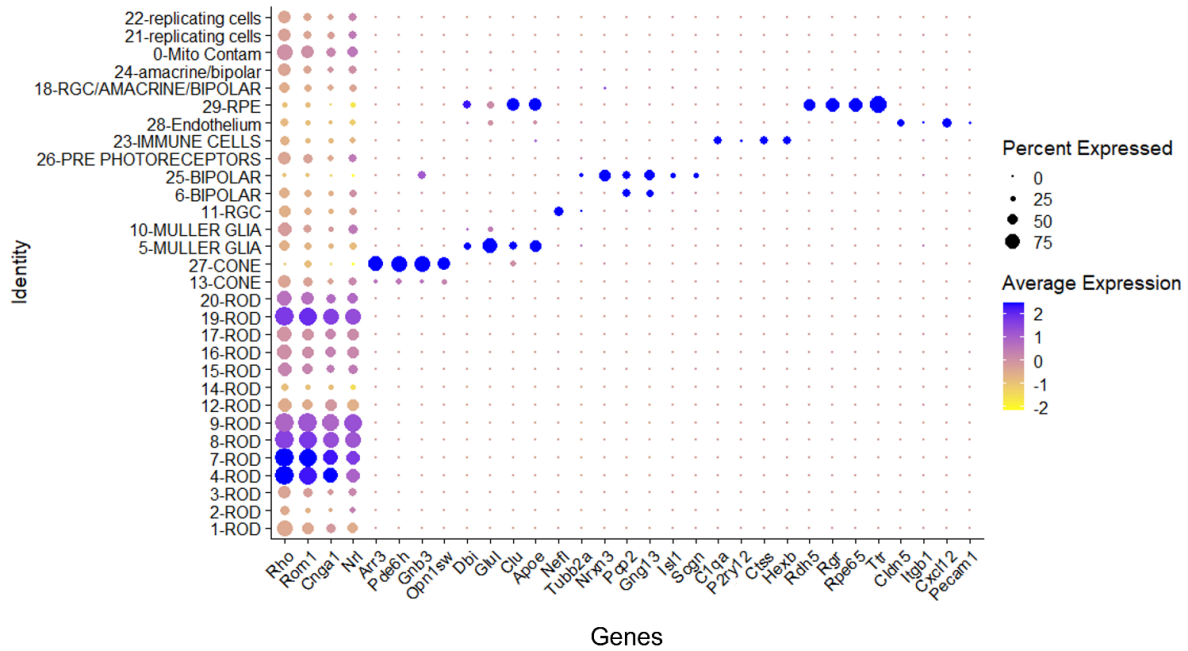


Fig4.3.3.2.4 Dot plot of the integrated P2Y6^{-/-} experiment samples, showcasing expression of differential cell type markers in different clusters.

| Samples | total | No. Rod | No. Cone | No. MG | No. Bipolar | No. RGC | No. Immune cells | No. RPE | No. Endothelium | No. Others |
|---------|--------|---------|----------|--------|-------------|---------|------------------|---------|-----------------|------------|
| WT1 | 20955 | 11821 | 827 | 2035 | 957 | 792 | 138 | 24 | 48 | 4313 |
| WT2 | 12595 | 9035 | 364 | 628 | 397 | 259 | 78 | 12 | 24 | 1798 |
| WT3 | 22681 | 12111 | 985 | 4258 | 1197 | 886 | 151 | 3 | 51 | 3039 |
| KO1 | 18132 | 9988 | 864 | 1871 | 1101 | 1110 | 93 | 9 | 20 | 3076 |
| KO2 | 8859 | 6390 | 300 | 613 | 334 | 273 | 56 | 7 | 6 | 880 |
| KO3 | 22550 | 14637 | 922 | 1626 | 2255 | 1099 | 218 | 5 | 6 | 1782 |
| Total | 105772 | 63982 | 4262 | 11031 | 6241 | 4419 | 734 | 60 | 155 | 14888 |

Table4.3.3.2.1 Cell counts of different cell types from P2Y6^{-/-} experiment samples. “Others” include cell types such as amacrine, progenitor, or empty cell clusters which were not the focus of the study.

| Samples | % Rod | % Cone | % MG | % Bipolar | % RGC | % Immune cells | % RPE | % Endothelium | % Others |
|---------|--------|--------|--------|-----------|-------|----------------|-------|---------------|----------|
| WT1 | 56.411 | 3.947 | 9.711 | 4.567 | 3.780 | 0.659 | 0.115 | 0.229 | 20.582 |
| WT2 | 71.735 | 2.890 | 4.986 | 3.152 | 2.056 | 0.619 | 0.095 | 0.191 | 14.276 |
| WT3 | 53.397 | 4.343 | 18.773 | 5.278 | 3.906 | 0.666 | 0.013 | 0.225 | 13.399 |
| KO1 | 55.085 | 4.765 | 10.319 | 6.072 | 6.122 | 0.513 | 0.050 | 0.110 | 16.964 |
| KO2 | 72.130 | 3.386 | 6.920 | 3.770 | 3.082 | 0.632 | 0.079 | 0.068 | 9.933 |
| KO3 | 64.909 | 4.089 | 7.211 | 10.000 | 4.874 | 0.967 | 0.022 | 0.027 | 7.902 |
| Total | 60.490 | 4.029 | 10.429 | 5.900 | 4.178 | 0.694 | 0.057 | 0.147 | 14.076 |

Table4.3.3.2.2 Cell count distribution of different cell types from P2Y6^{-/-} experiment samples.

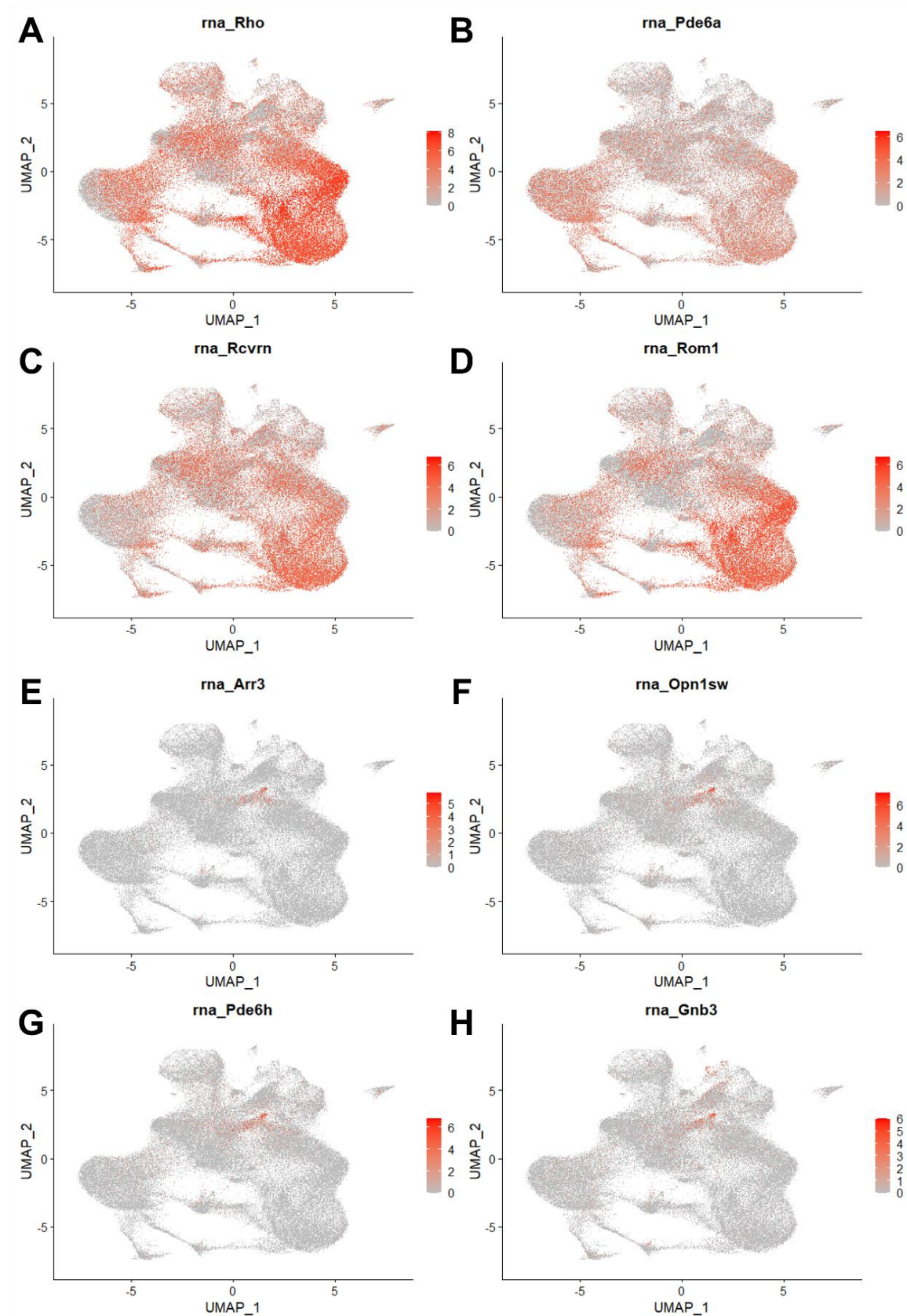


Fig4.3.3.2.5A-D: Rod cell markers expression. E-H: Cone cells markers expression.

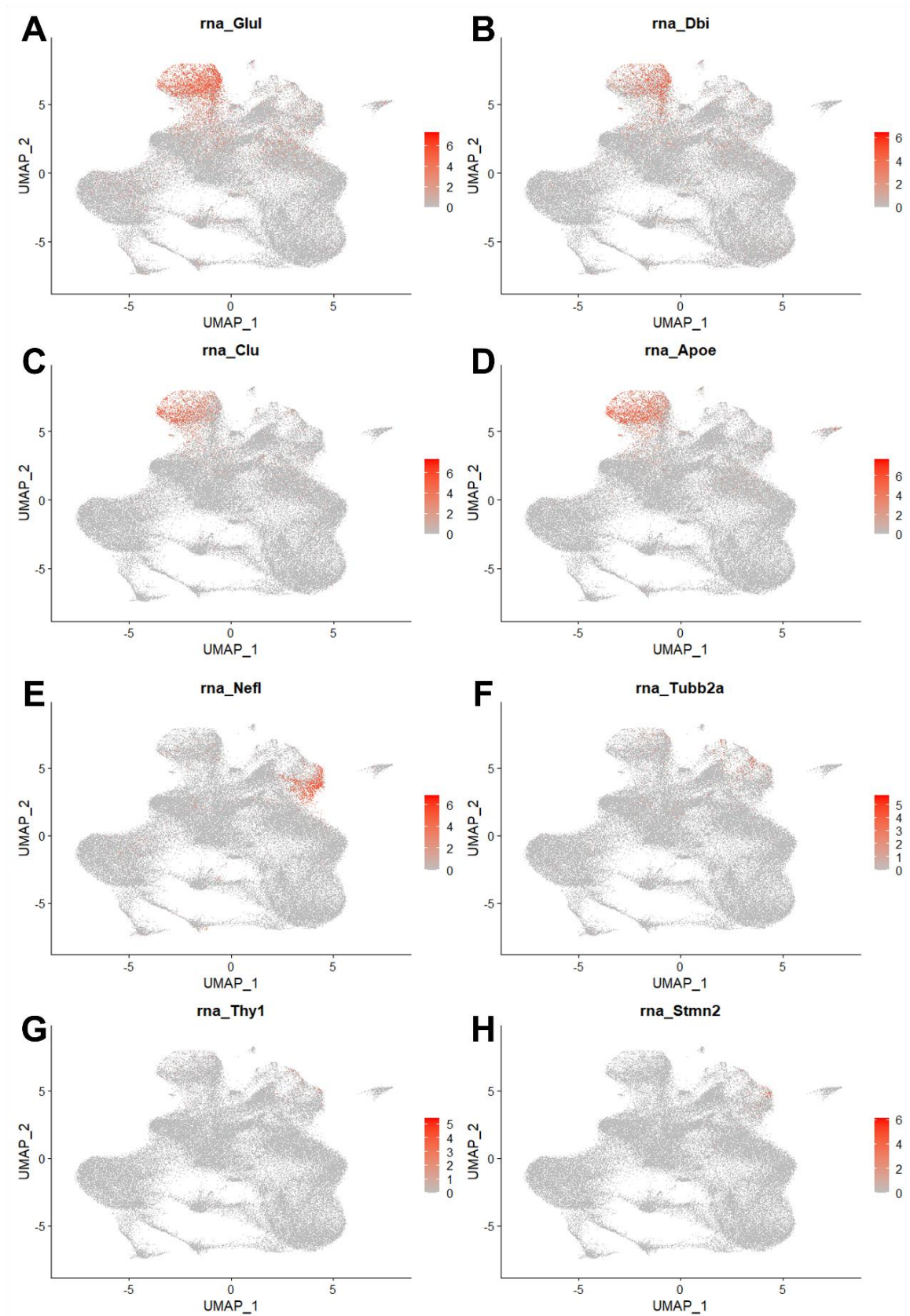


Fig4.3.3.2.6A-D: Müller glia markers expression. E-H: RGC cell markers expression.

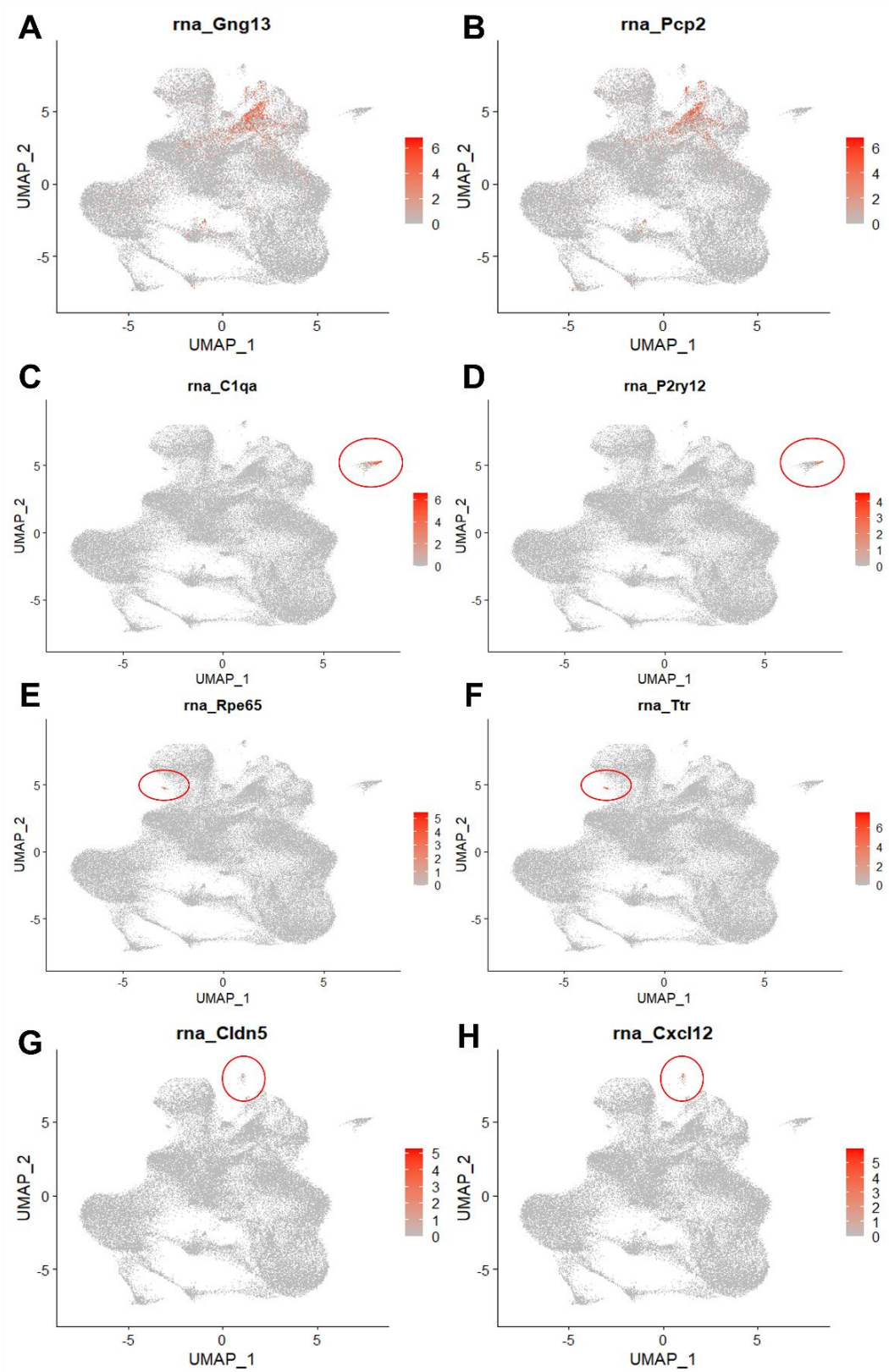


Fig4.3.3.2.7A-B: Bipolar cell markers expression. C-D: Immune cell markers expression. E-F RPE cell markers expression. G-H Endothelium cell markers expression

4.3.3.3 Subclustering of RGCs

With the *ABCA1*^{-/-} sample, numbers of RGCs were extremely limited due to the small number of cells included per sample, and that the population of RGCs within the retina is much lower in comparison to other cell types. As our studies focuses on studying glaucoma which primarily affects RGCs, maximizing the number of RGCs that can be found from the samples were important. Fig.4.3.3.3.1 showed violin plots of cells that expressed RGC cell markers *Nefl*, *Thy1*, *Stmn2*, *Resp18*. Cells expressing these genes were found mostly in cluster 11, hence labelled as the “RGC” cluster, but some were also found in cluster 18. As the lists of differentially expressed genes of cluster 18 included markers of RGCs, amacrine cells and bipolar cells, the identity of this cluster cannot be precisely defined. To prevent ignoring cells that may be RGCs that were not included in cluster 11, cells from both cluster 11 and cluster 18 were separated to create their own subset of data.

Subset data from cluster 11 and 18 were combined and processed through the *Seurat* analysis pipeline. This includes PCA, where the first 15 PCs were used for clustering using a granularity resolution of 0.7, and the results were projected as UMAP as shown in Fig4.3.3.3.2. Expression of RGC markers were once again inspected, and those with higher visibility were shown in Fig4.3.3.3A-D. These figures suggested the subclusters of cells in the middle expressed high levels of RGC markers, thus subcluster 6 was labelled as RGCs. Fig 4.3.3.3.2E showed the distribution of subcluster 6 and labelled with WT or KO sample origin. A noticeable small cluster of cells below the majority of the RGCs were shown to be only originating from KO samples, suggesting those RGCs might have KO specific expression and behaviour which differs from that from WT samples.

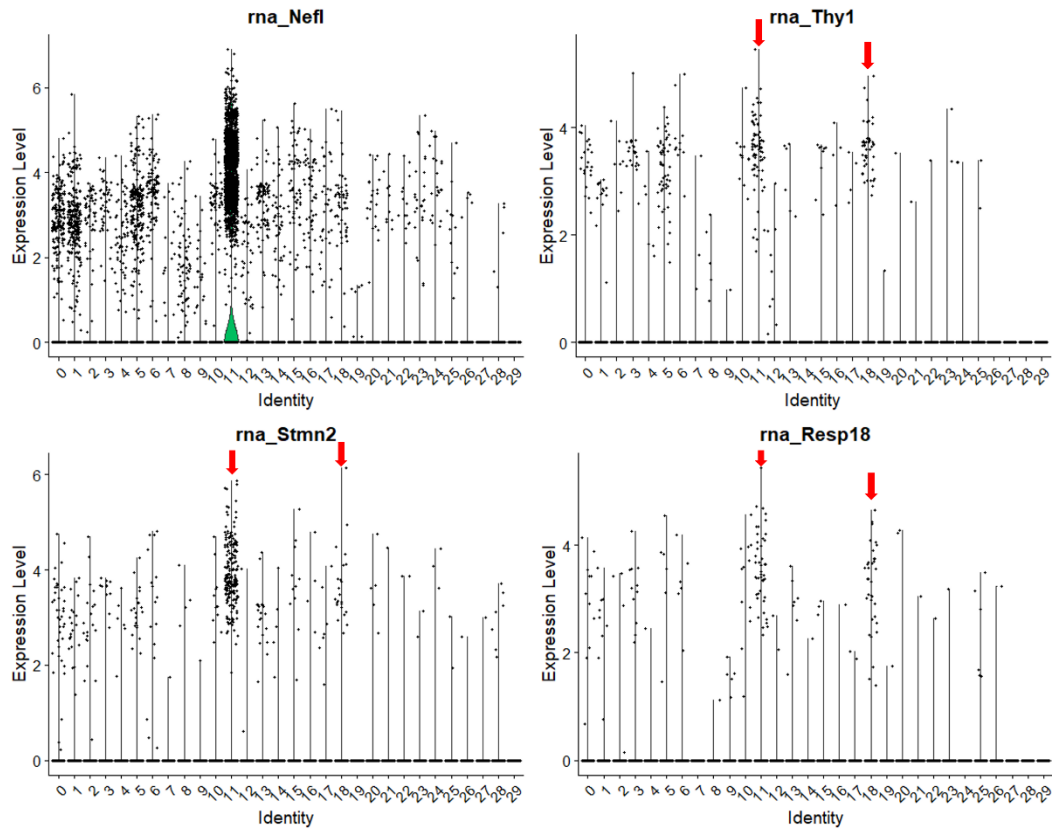


Fig4.3.3.3.1 RGC cell marker expression plot of integrated dataset. High level of expression was found in cluster 11, and some in cluster 18.

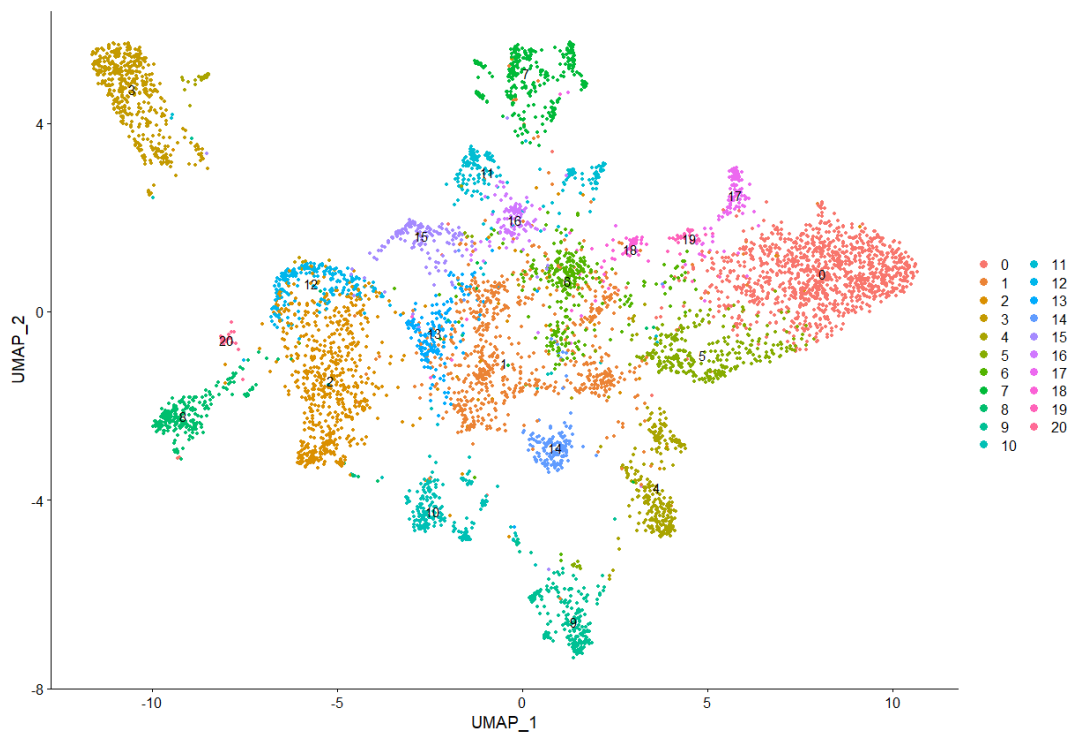


Fig4.3.3.3.2 UMAP Subclustering of cells from cluster 11&18.

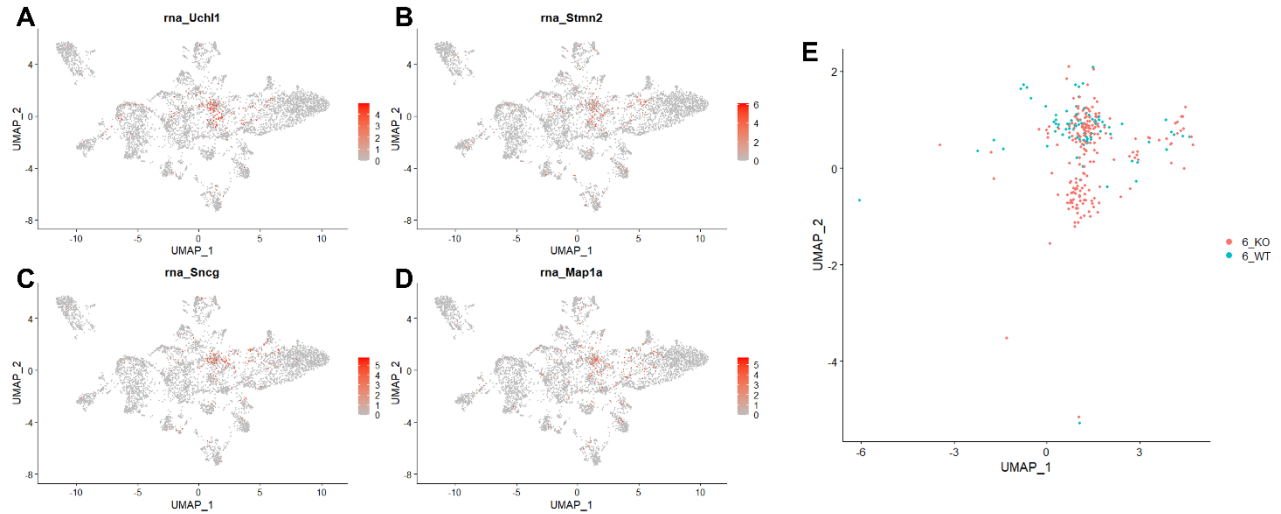


Fig4.3.3.3A-D: RGC cell marker expression in subclustering plot. E: Subcluster 6 split by WT and KO.

4.3.3.4 Subclustering of Glial cells

Identification of Müller glia and astrocytes was attempted using a similar approach. As mentioned previously in Chapter 3, Müller glia and astrocytes share common markers, thus the clusters were labelled as “Müller glia” in Fig4.3.3.2.3 may also contain astrocytes, in which a more detailed subclustering is required to refine the labelling of these glial cell types.

Fig4.3.3.2.6A-D showed high level of expression of some markers in cluster 5, while a smaller amount was also found expressed in cluster 10. Again, to maximize the number of glial cells included, cells from cluster 5 and 10 were combined and proceeded with subsequent subclustering analysis. Fig4.3.3.4.1 showed the UMAP projection created after clustering with a granularity resolution of 0.4 using the top 20 PCs from the PCA.

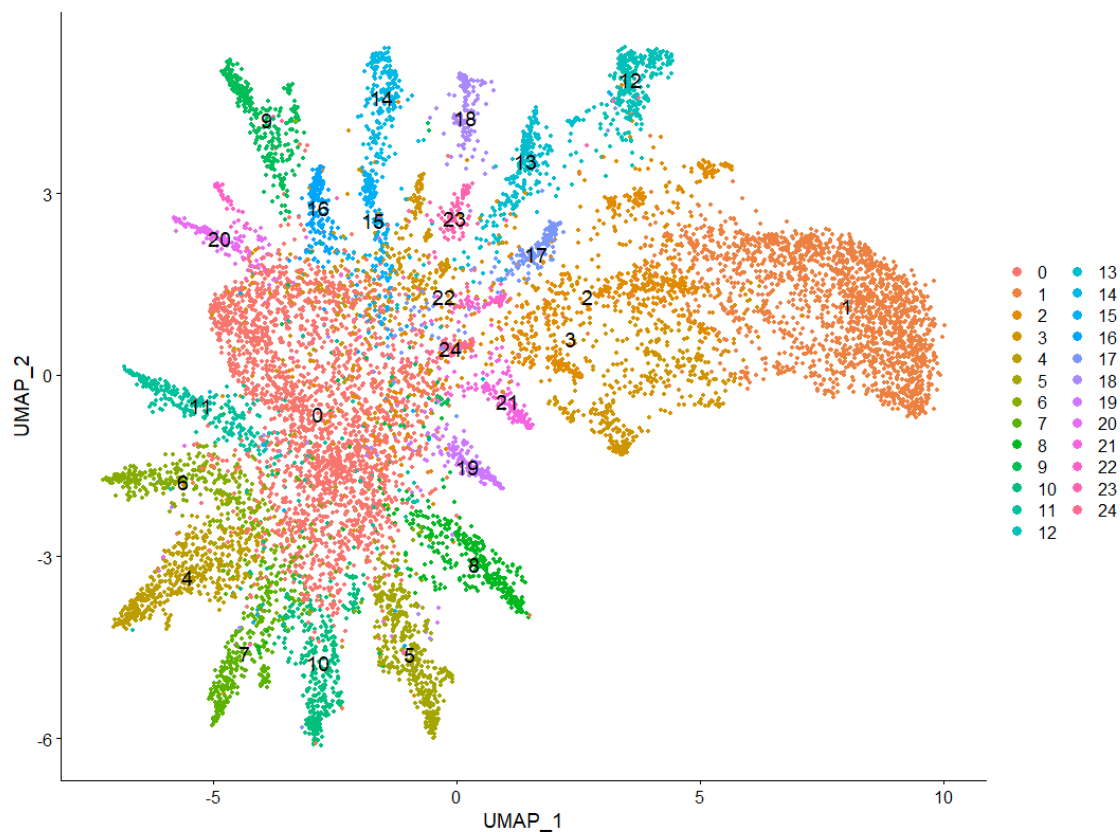


Fig4.3.3.4.1 UMAP subclustering of cells from cluster 5 & 10.

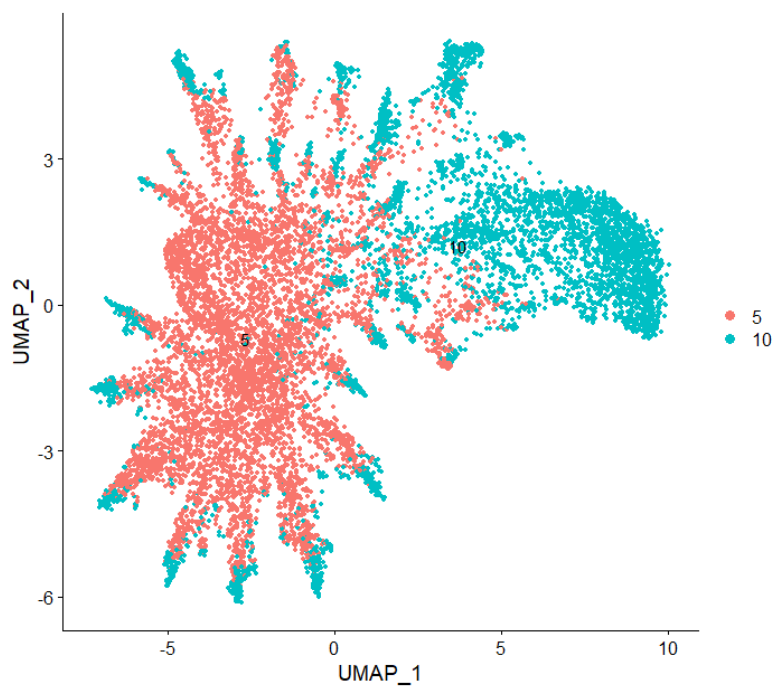


Fig4.3.3.4.2 Müller glia subcluster UMAP split by cluster origin.

Fig4.3.3.4.2 showed the distribution of cells from cluster 5 and 10 of the original clustering. The plot showed that majority of cells from cluster 5 belongs to the left area, while most of the cells from cluster 10 belongs to the right region. To identify cluster and cell type markers, differential expression of each cluster compared against the rest of the cells were obtained using *FindMarker* function. Expression of different Müller glia or astrocyte markers are displayed in Fig4.3.3.4.3A-I, with Fig4.3.3.4.3J showing the dot plot of genes from the top 2 most expressed genes from each subcluster. Subcluster 1 and 2, which are predominantly made up of cells from cluster 10, expressed high levels of rod photoreceptor markers such as *Pde6g* and *Rho*, and are thus not included in the Müller glia or astrocyte label. Majority of the rest of the cells expressed some levels of *Glul*, *Clu*, and *Apoe*, which are known Müller glia or astrocyte markers. Interestingly, almost all other subclusters exclusively highly expresses certain Müller glia markers, such as *Spc25*, *Vim*, *Sparc*, *Gnai2*, *S100a16* or astrocyte marker *Aqp4*. As there is a lack of combinational markers that indicate specific subclusters to be expressing multiple astrocyte markers, isolating astrocytes from the Müller glia population can be difficult without clear exclusive markers. Thus although subcluster 20 exclusively express high level of *Aqp4*, further verification is needed to label such cluster as astrocytes deterministically.

4.4 Ontological Analysis

Brief ontological analysis was then performed using differential expression gene lists obtained from three comparisons: “ALL” category labelled comparison between all cells from KO samples and that of WT samples, “MG” category labelled comparison between all Glial cells (classified in chapter 4.3.3.3) from KO and WT samples, and “RGC” labelled such comparison of RGCs from chapter 4.3.3.2. Three differential expression gene lists were inputted into IPA, providing data about the canonical pathways that were activated. Resultant canonical pathways lists were compared across three categories and vertically combined.

After inspection of the resultant table, pathways related to 4 biological events are listed here. “Neuroinflammation” and “Senescence” related pathways were displayed in Table 4.4.1 and Table 4.4.2, sorted by descending z-score of the comparison between all cells (ALL). Activation of these pathways were found to be high in “ALL” and “MG” category, signifying signs of neuroinflammation and senescence in most KO samples, including Müller glia. The activation of neuroinflammation were limited in the RGC clusters, while strong activation of senescence related pathways such as “EIF2 signaling”, “SNARE signaling” and “Ephrin Receptor signaling” were found with high confidence.

“Angiogenesis and Hypoxia” and “Fibrosis” related pathways were displayed in Table 4.4.3 and Table 4.4.4, sorted by the descending z-score of the comparison between all Müller glia (MG), as these events have been related to the roles of Müller glia and astrocytes in the retina under insults and injuries. Indeed, strong activation of pathways such as “FAK signaling”, “Signaling by Rho Family GTPases” and “HIF1 α Signaling” suggests angiogenesis and hypoxia related responses occurring in Müller glia, along with activation of “Pulmonary Fibrosis Idiopathic Signaling Pathway”, “Wound Healing Signaling Pathway” and “Hepatic Fibrosis Signaling Pathway” suggesting occurrence of fibrosis.

| Neuroinflammation | ALL | | MG | | RGC | |
|------------------------------------------------------------------------------|---------------|---------|---------------|---------|---------------|---------|
| | -log(p-value) | z-score | -log(p-value) | z-score | -log(p-value) | z-score |
| Neuropathic Pain Signaling In Dorsal Horn Neurons | 5.49 | 2.558 | 5.49 | -0.426 | 5.55 | 0.426 |
| Th2 Pathway | 1.94 | 2.53 | 1.94 | 1.897 | 1.67 | 0 |
| IL-3 Signaling | 0.652 | 2.121 | 0.654 | 1.414 | 0.666 | -0.707 |
| IL-17A Signaling in Gastric Cells | 1.42 | 2 | 1.42 | 2 | 1.43 | 1 |
| CCR3 Signaling in Eosinophils | 5.81 | 1.89 | 5.82 | 0.378 | 5.88 | 0.378 |
| B Cell Receptor Signaling | 0 | 1.886 | 0 | 0.943 | 0 | 0.728 |
| Regulation of IL-2 Expression in Activated and Anergic T Lymphocytes | 0 | 1.732 | 0 | 1.155 | 0 | 0.577 |
| Type I Diabetes Mellitus Signaling | 0 | 1.633 | 0 | 1.633 | 0 | -0.816 |
| Type II Diabetes Mellitus Signaling | 2.07 | 1.633 | 2.07 | 0.816 | 2.11 | -2.449 |
| IL-7 Signaling Pathway | 0.672 | 1.414 | 0.674 | 1.414 | 0.686 | 0 |
| CCR5 Signaling in Macrophages | 0 | 1.387 | 0 | 0.832 | 0 | -0.277 |
| B Cell Activating Factor Signaling | 1.03 | 1.342 | 1.03 | 1.342 | 1.05 | -0.447 |
| IL-17 Signaling | 0 | 1.155 | 0 | 1.732 | 0 | 0.577 |
| Osteoarthritis Pathway | 6.68 | 0.898 | 6.68 | -0.539 | 6.34 | -0.898 |
| CD28 Signaling in T Helper Cells | 0 | 0.577 | 0 | 0 | 0 | 0 |
| IL-15 Production | 2.42 | 0.471 | 2.42 | 1.886 | 2.46 | -0.943 |
| Natural Killer Cell Signaling | 1.02 | 0.447 | 1.02 | 0.447 | 1.05 | 0.447 |
| Production of Nitric Oxide and Reactive Oxygen Species in Macrophages | 2.77 | 0.408 | 2.77 | 2.858 | 2.82 | -0.816 |
| Complement System | 3.61 | 0.378 | 3.61 | 0.378 | 2.97 | -1.134 |
| IL-13 Signaling Pathway | 0.635 | 0.302 | 0.636 | 1.508 | 0.65 | 0.302 |
| PKCθ Signaling in T Lymphocytes | 0 | 0.277 | 0 | 0.277 | 0 | 0 |
| MIF Regulation of Innate Immunity | 0.996 | 0 | 0.996 | 2.449 | 1.01 | 0 |
| Th1 Pathway | 0.547 | 0 | 0.391 | 0 | 0.272 | 0.378 |
| Neuroinflammation Signaling Pathway | 7.84 | -0.164 | 7.84 | 0.493 | 7.54 | 1 |
| Calcium-induced T Lymphocyte Apoptosis | 0 | -0.302 | 0 | -2.111 | 0 | 0.632 |
| Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses | 0.445 | -0.333 | 0.445 | -0.333 | 0.458 | -1 |
| Toll-like Receptor Signaling | 0.288 | -0.447 | 0.289 | 1.342 | 0.296 | 1.342 |
| T Cell Exhaustion Signaling Pathway | 0 | -0.632 | 0 | -1 | 0 | -1 |
| T Cell Receptor Signaling | 0 | -0.655 | 0 | 0.218 | 0 | -0.447 |
| Role of PI3K/AKT Signaling in the Pathogenesis of Influenza | 0 | -1 | 0 | 1 | 0 | -1 |
| Interferon Signaling | 1.84 | -1.134 | 1.84 | 1.134 | 1.86 | 0.378 |

Table4.4.1 Canonical pathways activation related to neuroinflammation found in “ALL”, “MG” and “RGC” KO cells expression profile.

| Senescence | ALL | | MG | | RGC | |
|----------------------------------------|---------------|---------|---------------|---------|---------------|---------|
| | -log(p-value) | z-score | -log(p-value) | z-score | -log(p-value) | z-score |
| Autophagy | 0.385 | 2.668 | 0.386 | 0.728 | 0.4 | 0.243 |
| JAK/STAT Signaling | 0.401 | 2.646 | 0.401 | 1.89 | 0.41 | -0.378 |
| SNARE Signaling Pathway | 3.03 | 2.4 | 3.03 | 0.218 | 3.08 | 2.4 |
| Sirtuin Signaling Pathway | 0 | 2.138 | 0 | 0.535 | 0 | 0 |
| NAD Signaling Pathway | 0.845 | 2.138 | 0.845 | 1.604 | 0.866 | -0.535 |
| BMP signaling pathway | 1.6 | 2.111 | 1.6 | 1.508 | 1.63 | 0.905 |
| IL-1 Signaling | 5.88 | 1.633 | 5.88 | 0.816 | 5.94 | 1.633 |
| Pyridoxal 5'-phosphate Salvage Pathway | 0.68 | 1.633 | 0.68 | 1.633 | 0.693 | -0.816 |
| ATM Signaling | 1.23 | 1.508 | 1.23 | 0.905 | 1.25 | 0.302 |
| CXCR4 Signaling | 8.27 | 1.225 | 8.28 | 1.633 | 8.37 | 1.225 |
| Chemokine Signaling | 4.24 | 1.213 | 4.24 | -0.243 | 4.28 | 0.243 |
| IL-6 Signaling | 0.836 | 1.155 | 0.836 | 1.732 | 0.854 | 0.577 |
| p70S6K Signaling | 0.58 | 1.155 | 0.582 | 1.155 | 0.595 | 0.577 |
| SAPK/JNK Signaling | 0 | 1 | 0 | 1.667 | 0 | 0.333 |
| ERB2-ERBB3 Signaling | 0.281 | 1 | 0.281 | 1 | 0.287 | 0 |
| TGF- β Signaling | 1.63 | 0.905 | 1.64 | 0.905 | 1.66 | 0.302 |
| mTOR Signaling | 1.89 | 0.905 | 1.89 | 2.111 | 1.93 | -0.905 |
| IL-8 Signaling | 7.65 | 0.898 | 7.66 | 2.694 | 7.27 | 0.365 |
| PI3K/AKT Signaling | 0 | 0.816 | 0 | 1.633 | 0 | 1.633 |
| AMPK Signaling | 1.66 | 0.632 | 1.67 | 1.897 | 1.71 | 1.265 |
| GADD45 Signaling | 1.52 | 0.333 | 1.52 | 0.333 | 1.54 | -0.333 |
| NF- κ B Signaling | 0 | 0.243 | 0 | 2.668 | 0 | -1.213 |
| Ephrin B Signaling | 5.5 | 0 | 5.5 | 0.632 | 5.55 | 0.632 |
| Senescence Pathway | 0.821 | 0 | 0.821 | -0.943 | 0.851 | 0 |
| Ephrin Receptor Signaling | 9.13 | -0.2 | 9.13 | -0.2 | 8.71 | 1.8 |
| p38 MAPK Signaling | 0.767 | -0.632 | 0.77 | 0 | 0.785 | 1.265 |
| Oxidative Phosphorylation | 1.49 | -1.604 | 1.49 | -1.604 | 1.51 | -1.604 |
| EIF2 Signaling | 3.31 | -1.807 | 3.31 | -1.291 | 3.37 | 2.324 |

Table4.4.2 Canonical pathways activation related to senescence found in “ALL”, “MG” and “RGC” KO cells expression profile.

| Angiogenesis and Hypoxia | ALL | | MG | | RGC | |
|-------------------------------------------------------------------------------|---------------|---------|---------------|---------|---------------|---------|
| | -log(p-value) | z-score | -log(p-value) | z-score | -log(p-value) | z-score |
| FAK Signaling | 1.04 | 1.706 | 1.04 | 3.411 | 0.924 | -0.216 |
| Signaling by Rho Family GTPases | 11.6 | 1.183 | 11.6 | 2.874 | 11.2 | 2.197 |
| Regulation Of The Epithelial Mesenchymal Transition By Growth Factors Pathway | 2.74 | 0.408 | 2.74 | 2.041 | 2.79 | 0 |
| HIF1 α Signaling | 1.99 | 1.4 | 1.99 | 1.8 | 2.03 | -0.2 |
| VEGF Signaling | 0.721 | 1 | 0.721 | 1.667 | 0.738 | 1 |
| Endothelin-1 Signaling | 4.56 | 0.928 | 4.56 | 1.3 | 4.63 | 0.186 |
| Regulation of Actin-based Motility by Rho | 4 | -0.775 | 4 | 1.291 | 3.59 | 1.807 |
| CDC42 Signaling | 0 | 0.775 | 0 | 1.291 | 0 | 1.291 |
| RAC Signaling | 1.61 | 0.632 | 1.61 | 1.265 | 1.36 | 2.53 |
| Integrin Signaling | 2.71 | -0.784 | 2.71 | 1.177 | 2.46 | -0.2 |
| Thrombopoietin Signaling | 0.75 | 1.89 | 0.75 | 1.134 | 0.762 | -0.378 |
| VEGF Family Ligand-Receptor Interactions | 0.788 | 1.134 | 0.79 | 1.134 | 0.804 | -0.378 |
| ILK Signaling | 1.66 | -0.218 | 1.67 | 1.091 | 1.7 | 0.218 |
| RHOA Signaling | 2.03 | 0.5 | 2.04 | 1 | 2.07 | 3 |
| Erythropoietin Signaling Pathway | 0.81 | 2.183 | 0.81 | 0.728 | 0.833 | 0.243 |
| Regulation Of The Epithelial Mesenchymal Transition In Development Pathway | 1.6 | 1.134 | 1.6 | 0.378 | 1.63 | 0.378 |
| WNT/ β -catenin Signaling | 5.05 | 0.784 | 5.05 | 0 | 5.12 | 0.784 |
| Inhibition of Angiogenesis by TSP1 | 3.94 | 0 | 3.95 | 0 | 3.97 | -0.707 |
| Apelin Endothelial Signaling Pathway | 7 | 1.213 | 7 | -0.243 | 7.08 | 0.243 |
| RHO GDI Signaling | 9.9 | -1 | 9.91 | -1 | 9.49 | -2.2 |

Table4.4.3 Canonical pathways activation related to angiogenesis and hypoxia found in “ALL”, “MG” and “RGC” KO cells expression profile.

| Fibrosis | ALL | | MG | | RGC | |
|-------------------------------------------------|---------------|---------|---------------|---------|---------------|---------|
| | -log(p-value) | z-score | -log(p-value) | z-score | -log(p-value) | z-score |
| Pulmonary Fibrosis Idiopathic Signaling Pathway | 7.42 | 0.98 | 7.43 | 2.941 | 7.54 | -0.14 |
| Wound Healing Signaling Pathway | 3.05 | 0.174 | 3.06 | 1.567 | 2.82 | -1.061 |
| Hepatic Fibrosis Signaling Pathway | 13.6 | 0.64 | 13.6 | 1.152 | 13.3 | -0.128 |
| Inhibition of Matrix Metalloproteases | 0.81 | -2 | 0.81 | 1 | 0.821 | -2 |
| Tumor Microenvironment Pathway | 1.4 | 0.447 | 1.4 | 0.894 | 1.43 | 1.789 |
| Apelin Cardiac Fibroblast Signaling Pathway | 1.63 | -1.342 | 1.63 | -1.342 | 1.64 | 0.447 |

Table4.4.4 Canonical pathways activation related to fibrosis found in “ALL”, “MG” and “RGC” KO cells expression profile.

4.5 Discussion

4.5.1 Sample quality of single cell suspension

Sample preparation materials and protocols were different for *P2Y6*^{-/-} experiment in comparison to *ABCA1*^{-/-} experiment, as the procedures were conducted during the pandemic where there was limited access to different kits and solutions. Our collaborators in Japan handled the samples and employed a different dissociation protocol, which unfortunately yielded less than satisfactory cell viability samples. Sample qualities were also varied among samples as 3 sets of experiments were carried out separately for 3 pair of retinal tissues, causing the original estimation of cell number to range between 11,443-49,480 cells per sample. Auto-estimation of the contamination fraction by *SoupX* suggested most of the cells contained over 60-90% contaminated expression, implying to distinguish between actual cell expression and contamination expression can be very difficult.

This data demonstrates the importance of refined sample preparation procedures, ensuring the condition and viability of tissue samples have to be kept at high standards for drop-seq based single cell RNA-seq experiments. Computational algorithms developed by bioinformaticians can provide extensive assistance in correcting the artifacts and estimating the actual expression profiles, as demonstrated by *SoupX* for this dataset, whereby expression of background contamination profile were removed and allowed distinctive clusters to form more separately. However, these post-sequencing analysis methods rely on statistical assumptions, which may not always accurately depict the actual single cell expression profile, thus a higher quality sample preparation procedure should be implemented to eliminate the needs for downstream data correction. Different levels of parameters used in filtering and correction (such as number of cells per sample, contamination fraction, number of reads or genes per cells, and percentage of mitochondrial genes) should also be explored to construct a better understanding of the actual transcriptomic profile of the samples.

4.5.2 Potential subtypes of MGs

Chapter 4.3.3.4 explored the glial cells cluster and created subclustering plots for the data subset. Glial cells subclusters on the UMAP were distributed with a main core consist of mostly subcluster 0, while other subclusters expands outwards towards different directions. With accordance to the dot plot in Fig.4.3.3.4J, subcluster 0 expresses different glial markers at fairly low level but with high percentage distributions among all cells within the subcluster, while other subclusters expressed strong and high percentage expression of a specific marker gene. This phenomenon observed in the UMAP plot may have two interpretations.

Firstly, the occurrence of these subclusters may be due to gene dropout phenomenon observed in Drop-seq single cell RNA-sequencing. Subcluster 0 may represent cells where their expression of glial cell markers was captured and sequenced, while cells from other subclusters did not have expression of other glial makers detected due to low sequencing depth, or “gene-dropout”. Cells from these subclusters would have been projected closer to the central subcluster 0 if expression of other glial cell markers were detected during sequencing, making them less unique and more similar to other glial cells. Alternatively, this projection occurred due to actual expression difference within the same cell type, in which the clustering and UMAP algorithm forced them apart by maximizing the variation between cells. This implies the separation of different subclusters reflects the biological difference in expression profiles, thus each subcluster represent a subtype of glial cells which may have different functions, behaviour or morphology, but most importantly their response to injury.

In fact, the reality represented by this data may be a mixture of both hypotheses, where the differences between subclusters are due to both biological difference and sequencing artifacts. Marker genes identified with high expression level for each subcluster may provide clues as to the major function and

roles of the subtype of glial cells in the retina, and their differential expression in KO samples may provide insight to the subtype specific response to IOP injury that contributes to glaucoma phenotype.

4.6 Conclusion

In this chapter, we have implemented and performed the single cell RNA-seq analysis experiment on 3 WT and 3 *P2Y6*^{-/-} mouse retinal samples. Limited viability samples prepared by our collaborators provided sequencing data with slightly higher background expression contamination, thus the *SoupX* algorithm package was implemented to estimate contamination level and successfully removed and corrected the expression profile of the samples. Full *Seurat Integration* pipeline was implemented and identified different cell type clusters. RGCs and glial cells were selected for subclustering, in which 22 glial cells subclusters provided unique subtype markers. Ontological analysis suggested activation of neuroinflammation and senescence throughout the KO retina, as well as the glial cell population. Additionally, KO glial cells shows activation of pathways related to angiogenesis, hypoxia and fibrosis, suggesting its response to injury due to high IOP environment caused by the gene KO.

Chapter 5: Discussion

5.1 Preface

The experiments conducted in this project aimed at understanding the pathogenesis and development of glaucoma, through the use of single cell transcriptomic analysis on two different glaucoma mouse models. This final chapter summarizes the key findings of the experiments, discuss the impact and potentials of the analysis, account for limitations and outline future work to build on the findings of this research.

5.2 Key findings

5.2.1 Chapter 2: *ABCA1*^{-/-} glaucoma mouse model

ABCA1 gene knockout mouse model was used in this experiment as a model for Normal Tension Glaucoma (NTG). 3 wild type (WT) and 5 knockout (KO) mouse retinal samples were used to perform 10X Genomics single cell RNA-seq analysis, where both wet and dry lab pipelines were implemented. Large single cell sequencing data was analysed using *Seurat* pipeline, creating well-integrated data with visualization and clustering to identify different retinal cell types. Ontological analysis was performed by comparing WT and KO samples from all cells, Müller glia cluster, astrocyte cluster, retinal ganglion cell cluster and immune cell cluster. Neuroinflammation and senescence pathways were activated in KO samples, with specific mechanisms identified for different cell types. Activated microglia were identified in the immune cell cluster of the KO samples, suggesting a possible invasion of microglia as a form of neuroinflammatory response. Neurodegeneration pathways were found to be activated in RGCs clusters, in particularly a subtype of RGCs expressing genes enriched in the glutamatergic synapse pathway being more vulnerable to injury induced by the effects of the KO.

5.2.2 Chapter 3: *P2Y6*^{-/-} glaucoma mouse model

P2Y6 gene knockout mouse model was used in this experiment as a model for high Intraocular pressure (IOP) glaucoma mouse model. Three WT and three *P2Y6* KO mouse retinal samples were prepared and

performed with 10X Genomics single cell RNA-seq experiment by our collaborators. Cell dissociation and sample preparation was performed with challenges, leading to a low viability sample with highly contaminated sequencing data. *SoupX* package was first implemented onto each individual samples to identify background contamination percentage and profile, to correct expression and remove noisy contamination. Huge dataset with high numbers of cells per sample were then integrated using *Seurat's* integration pipeline after filtering, and implemented clustering algorithms to identify specific cell types such as RGCs and Müller glia. Subsequent subclustering were also performed, identifying RGCs accurately and suggested different Müller glia subclusters. Ontological analysis was performed by utilizing differential expression gene lists, where neuroinflammation and senescence were found activated when comparing all KO cells as well as Müller glia to their WT counterpart. Angiogenesis, hypoxia and fibrosis related pathways were identified to be strongly activated in KO Müller glia cluster, suggesting its activation response under the environmental stress and injury due to increase in IOP.

5.3 The power of single cell analysis

The data analysis of retinal tissues from two different glaucoma animal models positively showcased the power and impact of single cell transcriptomic analysis. The retina is a complex and heterogeneous tissue which performs the function of perceiving optical signals and act as the input for forming vision. This is achieved by the harmonious behaviour of different cell types within the retina to perform its tasks of perceive, transmit and integrate visual signals, but also maintaining homeostasis. For a complex disease such as glaucoma, pathogenesis and pathways of how it develops and progress is still difficult to study, due to the complicated nature of the retinal tissue. The deterioration of RGCs from glaucoma patient can originate from various different upstream sources, as different factors seem to cause different types of glaucoma as well as how their symptoms progress.

Single cell analysis conducted in this study provided extensive amount of details and data on the expression profiles of thousands of cells, and differences between WT and KO can be compared on different levels of clustering. Traditional bulk RNA-seq obtains average signals across all cell types within the retina, thus expression changes in cell types of small population will not be detected. As demonstrated by both experiments, the domination of rod photoreceptors population in the tissue implies differential expression signals obtained from bulk RNA-seq comparison between WT and KO would simply reflect changes in this cell type. For our study focus on the heterogenous nature of the glaucoma, constructing expression profile variations of multiple cell type which have a way smaller population, such as RGCs, glial cells or immune cells, is crucial in understanding their role and contributions to the progression of the disease.

The vast amount of single cell data also implies the potential of the dataset can be further explored in the future as our understanding of glaucoma or other retinal diseases progress. Although the focus of this study placed heavy emphasis on the aforementioned cell types, other cell types such as amacrine, bipolar, progenitors and endothelial cells were also included in these datasets. Single cell definition of this sequencing data implies the limit of our understanding of retinal cell types can be pushed further in the future, providing markers and expression data of different cell subtypes that may have different roles in the pathways that lead to glaucoma progression and development. Progression and development of bioinformatics tools can also contribute to discovering more potential on this dataset. *SoupX*, which is an actively updating opensource package, demonstrated its ability to correct contamination artifacts in the *P2Y6*^{-/-} experiment and obtained informative data about clusters of different cell types in the retinal samples. As the community and expertise of the field of single cell bioinformatics develop alongside computational methodologies, new ways of understanding and analysing the same dataset could provide more insight on to affected cell types, differential expressions and downstream affected ontological pathways.

5.4 Pathways activation in ABCA1^{-/-} induced glaucoma-like phenotype

Sequencing data from the *ABCA1*^{-/-} experiment strongly suggested high levels of neuroinflammation occurring throughout the retina. Ontological analysis showed related pathways activated in all cells, as well as particular pathways in Müller glia, astrocytes and retinal ganglion cells. Macrophage invasion evidence found in KO samples further suggests the recruitment of activated immune cell types, a strong indicator of neuroinflammatory response under the impact of the gene KO. These insults to the retina likely result in the degeneration and death of RGCs, leading to the glaucoma phenotype observed in this animal model.

To theorize the mechanism of how the gene KO leads to downstream occurrence of neuroinflammation, the actual function of the *ABCA1* protein can be further investigated. ATP binding cassette transporter A1 (*ABCA1*) is a transmembrane transporter protein, which is also known as cholesterol efflux regulatory protein (CERP). Mutation of this gene causes the Tangier disease, a genetic disorder characterized by a severe high-density lipoprotein (HDL) deficiency, due to the inability to secrete cholesterol to form such compound (Oram, 2000). Multitudes of other diseases related to inflammation such as asthma or atherosclerosis also found relation with *ABCA1* and its role of cholesterol homeostasis. (Chai et al., 2017; Soumian et al., 2005). In the context of glaucoma, observational studies have found patients with glaucoma had significantly higher total cholesterol levels than those without (Posch-Pertl, 2021), and administration of statin, an inhibitor of biosynthesis of isoprenoids including cholesterol, for five years or longer had a 21% lower risk of developing glaucoma (McGwin et al., 2004). Similar to other nervous tissues, astrocytes are suggested to be the site of cholesterol synthesis in the retina, thus it is evident that the KO of *ABCA1* in astrocytes causes disruption to cholesterol homeostasis for the whole retina environment.

To bridge the relation between cholesterol homeostasis disruption and neuroinflammation, we speculate it is due to the activation of senescence pathways. Cellular senescence is a stress response triggered by molecular damage, where cells stop dividing and enter a state of permanent growth arrest without cell death, sometimes leading to the senescence-associated secretory pathway that is associated with proinflammatory cytokine release (Birch & Gil, 2020). Senescence related pathways were found to be activated in KO samples, with specific pathways such as “Androgen signaling”, “Estrogen receptor signaling” and “mTOR signaling” being strongly activated in astrocyte clusters. In particular, studies have shown that lysosomal cholesterol activates the mTOR pathway, which is a key regulator of senescence. (Castellano et al., 2017; Xu et al., 2014). Cholesterol induced activation of senescence induces the release of SASP, thus causing neuroinflammatory phenotypes and neuronal damage. Limbad et al., 2020, reported astrocyte senescence promotes glutamate toxicity in cortical neurons, causing cognitive decline related to Alzheimer’s disease and related dementias. As for the retina, another study discovered linkage between *ABCA1* KO and senescence, whereby macrophage specific *ABCA1* KO resulted in accelerated senescence with accumulation of cholesterol in macrophages and promoted age-related macular degeneration (AMD) (Sene et al., 2013).

To summarize, it is possible that the impaired efflux of cholesterol due to the gene *ABCA1* KO causes cholesterol homeostasis interference, activating the mTOR pathway which promotes the SASP in senescent cells, inducing inflammatory response in the retina, leading to retinal ganglion cell damage and degenerative diseases such as glaucoma. Further lab experiments will need to be performed to evaluate cholesterol accumulation levels and senescence markers in KO retinal samples. If these results align with our hypothesis, it would be insightful to investigate on potential therapeutical solutions for this type of glaucoma by targeting either the cholesterol regulation pathway or senescence/ mTOR pathway. Statin, a drug to lower cholesterol which is typically known for reducing inflammation, have had suggestions in its protective effect against glaucoma development. (Xiao & Gong, 2017), possible due to its effect in

manipulating cholesterol levels in the retina. Rapamycin, an inhibitor of mTOR, have been shown to decrease aging and inhibit cellular senescence (R. Wang et al., 2017). Experimental application of these medicines on *ABCA1* KO mice would further elicit the details and efficacy on manipulating these pathways to control glaucoma pathogenesis.

It is important to acknowledge that while animal studies offer valuable insights, the translation of findings from mouse models to human clinical applications is a complex process. Glaucoma, being a multifaceted disease with diverse subtypes and underlying pathways, presents challenges in uncovering mechanisms contributing to its development across species. In the study by Hu et al. (Chunchun Hu, 2020), the role of *ABCA1* in trabecular meshwork tissue was explored, where it governed the permeability of endothelial cells within Schlemm's canal. This regulation impacted transendothelial electrical resistance and exhibited a potential to reduce intraocular pressure (IOP) by facilitating drainage. It should be noted that our research study did not include analysis of the trabecular meshwork or Schlemm's canal tissues, thus limiting direct comparisons regarding the involvement of *ABCA1* in mouse disease contexts. Moreover, a crucial aspect for exploration lies in examining cell-type-specific expression patterns within human tissues, including both drainage tissues and retinal neurons. Such experiments would facilitate discovering similarities and disparities between mouse and human samples. While previous studies have pursued various disease focuses, contextualizing and substantiating the role of *ABCA1* in cholesterol homeostasis within human astrocytes, particularly within the context of glaucoma, will be essential for the success of discovering effective and applicable potential therapeutic targets in the future.

5.5 Pathways activation in *P2Y6*^{-/-} induced glaucoma-like phenotype

Animal experiment data of aged *P2Y6*^{-/-} mouse showed increase in IOP, leading to RGC death and degeneration, representative of the pathogenic phenotype from high IOP related glaucoma. Single cell ontological analysis data obtained from this study suggested neuroinflammation and senescence

pathways being activated when comparing all cells from KO with WT samples, as well as the Müller glia cluster. Similar to *ABCA1*^{-/-} experiment, these mechanisms may be involved in contributing to the deterioration of the overall retina health and degeneration of RGCs.

In addition, strong angiogenesis, hypoxia, and fibrosis related pathways were found activated in the Müller glia cluster, as suggested by Table 4.4.3 and 4.4.4. Studies have shown the roles of Müller glia include the regulation of neuronal function and blood vessels formation, and its response to stress and injury inducing the upregulation of inflammatory responses and activating angiogenesis in the later stages of retinal diseases (Fletcher et al., 2021; Hu et al., 2014). Others have compared Müller glia to the fibroblasts within neuronal tissues which mediates the tissue-healing responses and activate fibrosis (Bringmann & Reichenbach, 2001; Friedlander, 2007; Roy et al., 2016). The response of glial cells to environmental stress, such as increase in IOP, is likely to trigger neurodegenerative processes such as gliosis, releasing pro-inflammatory cytokines, or even forming scar-like tissues which could lead to irreversible damage to the retina, especially the RGCs. Further detailed studies on how these events get triggered and progress could potentially uncover therapeutical pathways and targets in which glaucoma phenotypes can be inhibited or controlled. Utilizing subtypes of glial cells identified in this study would also benefit in identifying markers of glial cell sub-population responsible for reactive behaviour and phenotypes.

Angiogenesis also plays a huge role in controlling IOP through aqueous humor outflow regulation. Researchers have shown that Vascular endothelial growth factor (VEGF), a crucial signalling protein for the process of angiogenesis, are secreted by trabecular meshwork cells in response to IOP-dependant mechanical cues, which modulates the permeability of the endothelial cells of Schlemm's canal via VEGF receptor-2 (VEGFR-2) (Fujimoto, et al., 2015) (Ester Reina-Torres, 2017). The permeability or hydraulic conductivity of the Schlemm's canal inner walls thus regulates aqueous humor outflow and IOP homeostasis. Patients administered with anti-VEGF injections were found to have significant reduction in

outflow facility and elevated IOP (Wen, et al., 2017). P2Y6 gene belongs to the P2y receptors family, which can be found in trabecular meshwork and ciliary body tissues as well. More interestingly, some studies have suggested P2y receptor to be involved in angiogenesis via VEGFR-2 signalling in human cardiac endothelial cells (Sharif M. Rumjahn, 2007) (Martinez, 2022). The VEGF and angiogenesis signalling pathway disturbed as shown in our dataset from P2Y6^{-/-} mouse tissues suggest a potential angle to approach understanding P2y receptor family's role in IOP regulation, thus by conducting single cell experiment on the tissues involved in the aqueous outflow to identify transcriptomic changes in trabecular meshwork tissues and Schlemm's canal endothelial cells, will provide much more insight and value to the area of research.

Specific neurodegenerative pathways activation in RGCs were also suggested by Table 4.4.3. "RhoA Signaling" was highly associated with degenerative disease such as Parkinson's, Alzheimer's and Huntington's disease, through the regulation of mitochondrial homeostasis, autophagy and neuroinflammation (Schmidt et al., 2022). Meanwhile, Rac1 Signaling also activates autophagy in early stages of glaucoma which promotes RGC apoptosis (M. L. Zhang et al., 2020). Strong activation of these two pathways found in RGC cluster in our data could provide future leads onto the specific mechanisms in which RGCs react to injury and stressful environment, causing the deterioration and degradation resembled in glaucoma.

5.5 Limitations

5.5.1 Importance of sample preparation

Both *ABCA1*^{-/-} and *P2Y6*^{-/-} experiments provided significant and important results from the sequencing data analysis, but the process for quality control and contamination removal was more difficult for the latter. During the early stage of my studies, multiple iterations of attempts were made to refine the papain dissociation protocol for dissociating mouse retinal tissue into single cell suspension solution. This

included the concentration of enzymes, incubation time and temperature. Viability assessment was performed after different conditions until the optimal conditions were found. This protocol is then applied to the retinal tissue samples in the *ABCA1*^{-/-} experiment and resulted with over 90% viability samples for the 10X single cell experiment. Due to unfortunate circumstances with the pandemic conditions, limitations in animal transport internationally prohibited the single cell experiment for *P2Y6*^{-/-} samples to be processed in the UK. Dissociation of the samples were performed using a different protocol, and was performed with limited time and resources which yielded a much lower viability samples, thus more contaminated data samples. Although the subsequent analysis attempted to remove the contamination and correct expression profiles, a high viability and effective dissociation of the tissue is still crucial for the 10X Chromium system.

Thus, the refinement of the dissociated protocol and careful treatment of the dissociated samples, as well as confirmation on the viability of the samples before proceeding is a limiting factor for Drop-seq based single cell sequencing experiment such as that of 10X Genomics. There are however on-going developing methods which performs chemical fixation such as methanol fixation (Alles et al., 2017) and paraformaldehyde-fixation (Phan et al., 2021) which claims to be able to preserve RNA materials and fix cells or tissues and yield similar quality single cell sequencing samples. There are also new methods of performing single cell analysis which does not involve droplet-based technology, which allows for more flexible tolerance to different cell viability or fixation. Novel and cutting-edge technology could potentially assist with elevating limitations on sample preparation for single cell experiment in the future.

5.5.2 Rod photoreceptor filtering

Throughout the study, the issue with rod photoreceptor dominance in the retinal cell type population has been prominent. Although other cell types were identified and was able to be analysed and yield conclusive results, over 70% of the dataset was consisting of a cell type that was not as important as the

other ones in smaller population. Large number of droplets and subsequent sequencing resources were occupied by rod photoreceptors, which the other cell types could have benefited from by increasing the number of cells captured and the sequencing depth for other cell types of interest. Although the premise of a single cell experiment is to capture and identify all cell types within a tissue, for the context of the retina as demonstrated from our study, a pre-processing procedure during sample preparation to remove rod photoreceptors from the dissociated sample may be a beneficial act to consider. Fluorescence activated cell sorting (FACS) or magnetic-associated cell sorting (MACS) are filtering methods utilizing cell type surface marker to remove a particular subpopulation within a sample, and studies have been performed to isolate rod photoreceptors from adult mouse retina (Eberle et al., 2014; Feodorova et al., 2015). However, processes such as FACS or MACS will elongate the time for sample preparation, in which the cell viability of the samples may be affected and lead to more contamination from dead cell debris and RNA materials. Refinement of these protocols for the purpose of single cell analysis methods has to be performed to evaluate their efficiency and efficacy, and whether the loss in viability is worthwhile for a higher sample population of the non-rod cell types.

5.5.3 Gene dropout and determination of cell types

“Gene dropout” in single cell analysis describes the scenario when certain genes are detected in some cells but not the others, even if they are the same cell types and should be expressing similar genes. “Sequencing depth” or “sequencing saturation” describes the number of genes detected as percentage of the number of genes available for detection in the species. Although samples from *ABCA1*^{-/-} experiment has an average sequencing saturation at 84%, while that of *P2Y6*^{-/-} is at 90%, detection of the genes *ABCA1* and *P2Y6* in their respective WT samples were still difficult, despite verification by genotyping. This is likely because sequencing saturation can vary across cells within the sample, and gene dropout occur which doesn’t detect genes that were meant to be captured. Detection of other genes may not always be guaranteed as well, thus the expression of particular genes are usually inferred by similar

expression levels by other cells within a certain cluster. But for specific clusters with a smaller population, this can pose challenge in identifying their markers and make labelling difficult. Examples include glial cells in the retina, where Müller glia and astrocytes share multiple common markers while their specific unique markers are not well-defined. Thus, if certain cell type specific marker were not detected, such as *GFAP* for astrocytes, definitive separation between the two types can be difficult, as demonstrated from both experiments in this study.

Combinational markers were used in this study to consolidate the confidence in classifying certain cell clusters, which require prior knowledge to these markers from literature studies. Thus, it can be challenging to identify novel cell type and their respective markers by solely relying on single cell data clustering algorithms. Higher sequencing depth, which other advanced single cell experimental methods can achieve, would be beneficial for studies focusing on identifying novel cell types and markers. Otherwise, targeted gene expression single cell analysis methods may allow researchers to tag specific genes of interest such as well-defined cell markers, to ensure the expression of certain genes of interest can be measured, as well as facilitating the identification of specific niche cell types within the tissue.

5.6 Future work

5.6.1 Comparison between ABCA1^{-/-} and P2Y6^{-/-} model data

ABCA1^{-/-} and P2Y6^{-/-} mouse were established as mouse models for NTG and High IOP model glaucoma such as Primary open-angle glaucoma (POAG). As the single cell analysis experiment conducted in this study generated large amount of mouse sequencing data, it may be beneficial to compare the experimental across two experiments. WT data from both experiments contained expression profiles of different cell types within a normal mouse retina of similar age, thus these samples can be used as anchors for integration to minimize the batch effects across the two experimental conditions. All samples could then be corrected to integrate with each other, such that cell types across the two knockout conditions

can collapse into similar clusters, hence allowing differential expression profile comparison. Cell types and subtypes which shares expression pattern and geometric proximity in projections across knockout conditions may share commonly activated ontological pathways and functions, thus providing leads into the shared mechanisms that ultimately lead to the RGC degeneration observed in different types of glaucoma. On the contrary, cell types or expression patterns which differs between the two conditions may suggest their individual source and upstream targets of how their pathogenic phenotype originate. Common pathways and mechanisms are extremely useful in identifying potential therapeutical targets which stops the progression of different types of glaucoma and their phenotype, while the KO condition-specific pathways would indicate how each type of glaucoma initiate, thus suggesting the targets for preventative measures.

5.6.2 Validation of cell type and expression markers

Our study of the single cell data utilized clustering algorithms to determine clusters and label identities based on differentially expressed markers of each cluster. To relate these cell type markers to their physiological appearance and location in the tissue, immunostaining of these marker proteins will provide credibility and verification to these findings. For example, novel subtypes such as the RGC-1 subtype identified in the ABCA1^{-/-} experiment expressed certain markers which leads to prediction of their behaviour to be particularly vulnerable to excitotoxicity. By investigating the markers in the retinal tissue and identifying its physiological structure and the connections that this subtype of RGC have, further findings could be discovered and predict its role in glaucoma progression. Tran et al., 2019 performed a single cell profile study of just mouse RGCs and created a comprehensive molecular atlas of 46 RGC subtypes (Fig.5.6.2A), along with hierarchical cluster which predicts their function and morphological structures. An early attempt to integrate RGCs from ABCA1^{-/-} data samples into the cell atlas of their study is shown in Fig5.6.2B, seemingly suggesting it may belong to a particular subtype. This analysis may not be conclusive due to the limited number of RGCs found in our samples, but nonetheless suggest the

significance of validating the subtypes suggested in our data by referencing other studies or morphological staining. Similar analysis could be applied to astrocytes and Müller glia clusters, which could be helpful in consolidating exclusive specific markers for the two cell types, as well as differentiating subtypes of Müller glia suggested from the P2Y6^{-/-} datasets.

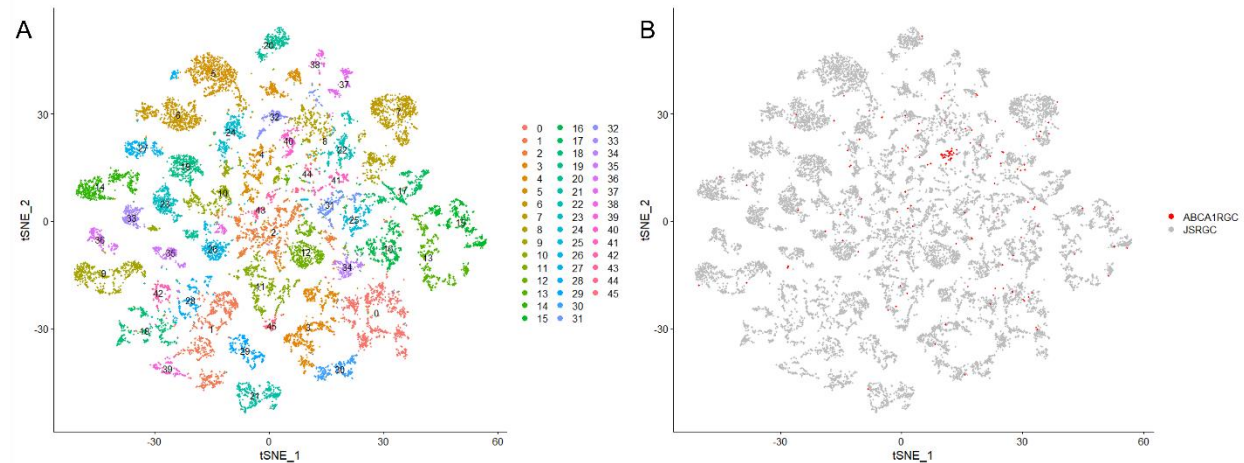


Fig5.6.2A: tSNE projection of the RGC atlas data from Tran et.al. study using Seurat pipeline. B: Integration of RGCs from ABCA1^{-/-} datasets into the RGC atlas.

Moreover, staining for markers found in immune cell types such as activated microglia or macrophage would provide strong evidence to support the findings of this study, identifying the patterns and regions in which neuroinflammation occurs. Quantification of gene expression related to neuroinflammation, senescence using methods such as quantitative polymerase chain reaction (qPCR) and cholesterol staining in KO samples could further provide insight onto the relationship between these mechanisms and glaucoma development.

5.6.3 Identification and application of potential drug targets

As suggested by the previous section 5.4, strong evidence of neuroinflammation occurring in ABCA1^{-/-} samples may be related to the disruption of cholesterol regulation in the retinal astrocytes, leading to senescence phenotype and release of SASP. Therapeutical drugs which target both cholesterol regulation or senescence could be applied onto experimental animal samples and observe their effect

on the glaucoma phenotype development. Application of these drugs into KO mouse samples should present an age-dependent effect in slowing down RGC degeneration. Single cell analysis of retinal samples obtained at different age between 3-12 months old, either with or without drug application, could provide insightful information on how glaucoma develops and progress, which cell types are involved at different stages, and what cell types and pathways do the drugs target.

5.6.4 Inclusion of other retinal structures

P2Y6^{-/-} experiment identified cell types and pathways affected in the retina in response to the effect of the gene knockout, which presented the phenotype of raised IOP. To complete the understanding of the initiation of such phenotype which leads to the degeneration of RGCs in the KO samples, inclusion of cell types and structures which control IOP and express P2 receptor family genes would be beneficial. Tissue structures near the base of the retina such as the trabecular meshwork, Schlemm's canal and the ciliary body (shown in Fig5.6.3) would benefit from single cell analysis by identifying gene expression changes in different cell types under KO conditions, which can provide insights into potential targets for preventive therapeutical solutions to prohibit glaucoma initiation.

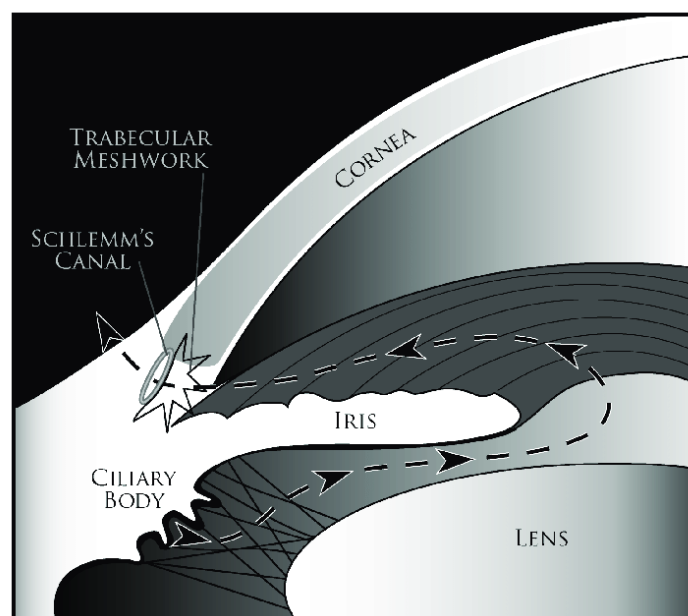


Fig5.6.3 : Schematic diagram illustrating the trabecular meshwork aqueous humor production and outflow pathway. (Goel, 2010)

5.6.5 Advanced single cell analysis techniques

Single cell analysis technology has been developing for the last decade, which is still a young and actively progressing field of study. Many variations on the technology continues to emerge in comparison to the options available during the beginning of this study. For example, split-pool ligation-based transcriptome sequencing (SPLiT-seq, X. Tang et al., 2019) utilizes combinatorial barcodes added onto well plates does not require formation of droplets to capture single cell like in Drop-seq. The commercialized option provided by Parse bioscience claims to sequence more cells and with higher sequencing depth (V. Tran et al., n.d.), with less contamination effects. Spatial transcriptomic is also an advanced single cell technology which allows the sequencing of single cells with conserved spatial information, allowing the reconstruction of expression profiles of different cell types found at their original location in the tissue. This technique could benefit the study of glaucoma mouse models as cell types (such as RGCs) have uneven distribution structurally on the retina (such as concentration on the optic nerve head), allowing easier identification of cell types but also observe structural changes/ damages and migration related phenomenon.

5.7 Conclusion

This study successfully analysed two gene knockout glaucoma mouse model retinal tissues using single cell RNA-sequencing analysis. We identified major evidence of neuroinflammation occurring in *ABCA1* KO samples, macrophage invasion, and discovered an RGC subtype that is particularly vulnerable. *P2Y6* KO samples showed activation of neuroinflammation and senescence throughout the retina, while KO glial population suggested upregulation of angiogenesis, hypoxia, and fibrosis related signaling pathways.

Our study demonstrated the effectiveness of utilizing single cell analysis to understand heterogeneous tissues and diseases that relate to them. This data contains extensive number of resources and potential leads for further discovery about the development and progression of glaucoma-like phenotype through

different pathways and cell types involved, and suggested potential mechanisms that could lead to further studies to develop therapeutical strategies in preventing, stopping or treating glaucoma.

Chapter X: Appendices

X.1 R code for Individual ABCA1-/- experiment samples: WT1 example

```
# Loading library -----
library(Seurat)
library(dplyr)
library(cowplot)

# Loading data -----
WT1.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/WT1/filtered_feature_bc_matrix/")
WT1 <- CreateSeuratObject(counts = WT1.data, min.cells = 3, min.features =
200, project = "10X_WT1", assay = "RNA")

# Finding percentage of mitochondrial gene expression -----
mito.genes <- grep(pattern = "^mt-", x = rownames(WT1@assays[["RNA"]]), value
= TRUE)
percent.mito <-
Matrix::colSums(WT1@assays[["RNA"]][mito.genes, ])/Matrix::colSums(WT1@assays
[["RNA"]])
WT1 <- AddMetaData(object = WT1, metadata = percent.mito, col.name =
"percent.mito")

#Violin plot: number of features, number of counts, percentage of mito genes
VlnPlot(object = WT1, features = c("nFeature_RNA", "nCount_RNA",
"percent.mito"), ncol = 3)
#Plotting features against each other
FeatureScatter(object = WT1, feature1 = "nFeature_RNA", feature2 =
"percent.mito")
FeatureScatter(object = WT1, feature1 = "nCount_RNA", feature2 =
"nFeature_RNA")

# Filter and Normalize -----
WT1 <- subset(x = WT1, subset = nFeature_RNA > 0 & nFeature_RNA < 2000 &
nCount_RNA <2000 & percent.mito > 0 & percent.mito < 0.5 )

#Normalization - Normalize by total expression, multiply by 10,000, log-
transform
WT1 <- NormalizeData(object = WT1, normalization.method = "LogNormalize",
scale.factor = 10000)

# Find Variable Feature -----
WT1 <- FindVariableFeatures(WT1, selection.method = "vst", nfeatures = 2000)
head(x = HVGInfo(object = WT1))

# Regressing unwanted variation -----
all.genes <- rownames(WT1)
WT1 <- ScaleData(WT1, features = all.genes)

# PCA -----
WT1 <- RunPCA(WT1, features = VariableFeatures(object = WT1))

# Clustering -----
WT1 <- FindNeighbors(WT1, dims = 1:15)
#a resolution parameter that sets the 'granularity' of the downstream
clustering, with increased values leading to a greater number of clusters. We
```

find that setting this parameter between 0.4-1.2 typically returns good results for single cell datasets of around 3K cells.

```
WT1 <- FindClusters(WT1, resolution = 1.5)
```

```
# TSNE -----  
WT1 <- RunTSNE(object = WT1, dims.use = 1:15, do.fast = TRUE, do.label=TRUE)  
# note that you can set do.label=T to help label individual clusters  
DimPlot(object = WT1, reduction = "tsne", label = TRUE, pt.size = 0.5)
```

```
# UMAP -----  
WT1 <- RunUMAP(object = WT1, dims = 1:15, do.fast = TRUE, do.label=TRUE)  
# note that you can set do.label=T to help label individual clusters  
DimPlot(object = WT1, reduction = "umap", label = TRUE, pt.size = 0.5)
```

```
# CellType-----  
-
```

```
FeaturePlot(object = WT1, features = c("Rho", "Sag", "Gnat1", "Pde6a")) #Rod
```

```
FeaturePlot(object = WT1, features = c("Arr3", "Gngt2", "Opn1sw", "Cngb3"))  
#Cone
```

```
FeaturePlot(object = WT1, features = c("Gap43", "Sncg", "Nrn1", "Thy1"))  
#RGC, Opn4, Pou4f1
```

```
FeaturePlot(object = WT1, features = c("Rlbp1", "Vim", "Glul", "Gpr37"))  
#MüllerGlia, Rgr, Sl00a1
```

```
FeaturePlot(object = WT1, features = c("Dab1", "Gad1", "Calb1", "Gad2"))  
#Amacrine
```

```
FeaturePlot(object = WT1, features = c("Calb1", "Onecut1", "Onecut2"))  
#Horizontal
```

```
FeaturePlot(object = WT1, features = c("Otx2", "Vsx2", "Prkca", "Isl1"))  
#Bipolar
```

```
FeaturePlot(object = WT1, features = c("Gfap", "Stat3", "Slc1a3", "Slc6a1"))  
#Astrocyte
```

```
# Non-Rod subset-----  
WT1_nonrod = subset(WT1, idents =  
c("10", "11", "7", "9", "6", "12", "14", "13", "0"))  
DimPlot(object = WT1_nonrod, reduction = "tsne")
```

X.2 R code of ABCA1-/- experiment samples integration

```
#Load in packages
library(Seurat)
library(ggplot2)
library(cowplot)
library(patchwork)

#Load in datasets
WT1.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/WT1/filtered_feature_bc_matrix/")
WT1 <- CreateSeuratObject(counts = WT1.data, min.cells = 3, min.features =
100, project = "10X_WT1", assay = "RNA")

WT2.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/WT2/filtered_feature_bc_matrix/")
WT2 <- CreateSeuratObject(counts = WT2.data, min.cells = 3, min.features =
100, project = "10X_WT2", assay = "RNA")

WT3.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/WT3/filtered_feature_bc_matrix/")
WT3 <- CreateSeuratObject(counts = WT3.data, min.cells = 3, min.features =
100, project = "10X_WT3", assay = "RNA")

KO1.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/KO1/filtered_feature_bc_matrix/")
KO1 <- CreateSeuratObject(counts = KO1.data, min.cells = 3, min.features =
100, project = "10X_KO1", assay = "RNA")

KO2.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/KO2/filtered_feature_bc_matrix/")
KO2 <- CreateSeuratObject(counts = KO2.data, min.cells = 3, min.features =
100, project = "10X_KO2", assay = "RNA")

KO3.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/KO3/filtered_feature_bc_matrix/")
KO3 <- CreateSeuratObject(counts = KO3.data, min.cells = 3, min.features =
100, project = "10X_KO3", assay = "RNA")

KO4.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/KO4/filtered_feature_bc_matrix/")
KO4 <- CreateSeuratObject(counts = KO4.data, min.cells = 3, min.features =
100, project = "10X_KO4", assay = "RNA")

KO5.data <- Read10X(data.dir = "C:/Alex/UCL-
Ophthalmology/SingleCellWork/ABCA1KO/KO5/filtered_feature_bc_matrix/")
KO5 <- CreateSeuratObject(counts = KO5.data, min.cells = 3, min.features =
100, project = "10X_KO5", assay = "RNA")

#Create list to loop through process of NORMALIZATION and FINDING VARIABLES
ALL.list <- c(WT1,WT2,WT3,KO1,KO3,KO4,KO5)

for (i in 1:length(ALL.list)) {
  ALL.list[[i]] <- NormalizeData(ALL.list[[i]], verbose = FALSE)
  ALL.list[[i]] <- FindVariableFeatures(ALL.list[[i]], selection.method =
"vst",nfeatures = 2000, verbose = FALSE)
}
```

```

#recommended dims 10-50
ALL.anchors <- FindIntegrationAnchors(object.list = ALL.list, dims = 1:30)

ALL.integrated <- IntegrateData(anchorset = ALL.anchors, dims = 1:30)

DefaultAssay(ALL.integrated) <- "integrated"

ALL.integrated <- ScaleData(ALL.integrated, verbose = FALSE)

#PCA
ALL.integrated <- RunPCA(ALL.integrated, npcs = 30, verbose = TRUE)
DimPlot(object = ALL.integrated, reduction = "pca")

#Plot to show variance represented by increasing PCA dimensions.
#Excess information included from increasing PCA dimensions decreases
exponentially
ElbowPlot(object = ALL.integrated)

#UMAP
ALL.integrated <- RunUMAP(ALL.integrated, reduction = "pca", dims = 1:20)

#Clustering Process
ALL.integrated <- FindNeighbors(ALL.integrated, dims = 1:20)
#a resolution parameter that sets the 'granularity' of the downstream
clustering, with increased values leading to a greater number of clusters. We
find that setting this parameter between 0.4-1.2 typically returns good
results for single cell datasets of around 3K cells.
ALL.integrated <- FindClusters(ALL.integrated, resolution = 1.5)

#Plot UMAP, grouped by original identity
DimPlot(ALL.integrated, reduction = "umap", label = FALSE, pt.size = 0.5,
group.by = "orig.ident")
#Plot UMAP, grouped by clustering algorithm
DimPlot(ALL.integrated, reduction = "umap", label = TRUE, pt.size = 0.5)

#Differential Expression, find markers in each cluster
ALL.integrated.markers <- FindAllMarkers(ALL.integrated, only.pos = TRUE,
min.pct = 0.25, logfc.threshold = 0.25)

#Export markers list as an Excel
write.csv(ALL.integrated.markers, "ALLint_ClustersRetinalBiomarkers.csv")

#Renaming clusters
new.cluster.ids <- c("0-Rod", "1-Rod", "2-Rod", "3-Rod", "4-Rod", "5-
(?) Horizontal/Amacrine", "6-Rod", "7-Rod", "8-(?) Pre-RGC", "9-(?) lens", "10-
(?) lens", "11-Bipolar (cone?)", "12-(?) lens", "13-Rod", "14-MüllerGlia", "15-
Contaminant", "16-Cone", "17-MüllerGlia", "18-Bipolar", "19-Bipolar Rod", "20-
Cone", "21-(?)", "22-(?) Horizontal/Amacrine", "23-Rod", "24-Bipolar Rod", "25-
MüllerGlia", "26-MüllerGlia/Astrocyte", "27-Microglia", "28-(?) Rod", "29-
Endothelial Cells", "30-Rod", "31-Retinal pigment epithelium")
names(new.cluster.ids) <- levels(ALL.integrated)
ALL.integrated.rename <- RenameIds(ALL.integrated, new.cluster.ids)

#Choosing specific clusters as subset for further investigation if needed
ALL.MG = subset(ALL.integrated, ids = c(14, 17, 25, 26))
ALL.mixedgroup = subset(ALL.integrated, ids = c(8, 21, 22))

```

```

#Plot using specific gene level expression
# CellType-----
-
FeaturePlot(object = ALL.integrated, features = c("rna_Rho", "rna_Sag",
"rna_Gnat1", "rna_Pde6a")) #Rod

FeaturePlot(object = ALL.integrated, features = c("rna_Arr3", "rna_Gngt2",
"rna_Opn1sw", "rna_Cngb3")) #Cone

FeaturePlot(object = KO.integrated, features = c("rna_Chra6",
"rna_Nrn1", "rna_Sncg",
"rna_Thy1", "rna_Resp18", "rna_Pou4f1", "rna_Slc17a6", "rna_Stmn2", "rna_Rbpms"))
#RGC, Opn4, Pou4f1

FeaturePlot(object = ALL.integrated, features = c("rna_Gpr37", "rna_Rgr",
"rna_Glul",
"rna_Vim", "rna_Rlbpl", "rna_Clu", "rna_Dkk3", "rna_Abca8a", "rna_Apoe"))
#MüllerGlia, Rgr, S100a1

FeaturePlot(object = ALL.integrated, features = c("rna_Tfap2b", "rna_Gad1",
"rna_Lrrn3", "rna_Gad2")) #Amacrine

FeaturePlot(object = ALL.integrated, features = c("rna_Calb1", "rna_Onecut1",
"rna_Onecut2")) #Horizontal

FeaturePlot(object = ALL.integrated, features = c("rna_Gng13", "rna_Vsx2",
"rna_Prkca", "rna_Isl1")) #Bipolar

FeaturePlot(object = ALL.integrated, features =
c("rna_Bhlhb4", "rna_Cabp5", "rna_Gabrr2", "rna_Hcn3", "rna_Isl1", "rna_Slc17a7"))
#Bipolar

FeaturePlot(object = ALL.integrated, features =
c("rna_Glrb", "rna_Gria2", "rna_Tacr3", "rna_Vsx1", "rna_Gja7")) #BipolarCone

FeaturePlot(object = ALL.integrated, features = c("rna_Rdh5", "rna_Rgr",
"rna_Rpe65", "rna_Rrh", "rna_Ttr")) #Bipolar

FeaturePlot(object = ALL.integrated, features =
c("Ndr2", "Aqp4", "Gfap", "S100a16")) #Astrocyte, data specific: S100a16, Espn,
Cd9, Hes1, Slc1a3

#Grouping WT and KO and find markers against each other
WT.integrated = subset(ALL.integrated, orig.ident == "10X_WT1"|orig.ident
=="10X_WT2"|orig.ident == "10X_WT3")
DimPlot(WT.integrated, reduction = "umap", label = TRUE, pt.size = 0.5)
Idents(WT.integrated) = "WT"

KO.integrated = subset(ALL.integrated, orig.ident == "10X_KO1"|orig.ident
=="10X_KO2"|orig.ident == "10X_KO3"|orig.ident == "10X_KO4"|orig.ident
=="10X_KO5")
DimPlot(KO.integrated, reduction = "umap", label = TRUE, pt.size = 0.5)
Idents(KO.integrated) = "KO"

```



```

ALL.combined <- merge(x= WT.integrated, y = KO.integrated)

ALL.combined.markers <- FindAllMarkers(ALL.combined, only.pos = FALSE,
min.pct = 0, logfc.threshold = 0)

write.csv(ALL.combined.markers, "ALLint_ALL_WTvsKO_DiffExpList.csv")

```

X.3 R code of SoupX analysis for P2Y6-/- experiment data

```

library(Seurat)
library(ggplot2)
library(SingleR)
library(dplyr)
library(cellidex)
library(RColorBrewer)
library(SingleCellExperiment)
library(SoupX)
memory.limit(size = 9e12)

#Setting directory, following script applies to 1st sample WT1
workdir = c("C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_WT1", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_WT2", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_WT3", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_KO1", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_KO2", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_KO3")
workdirvar = 1
setwd(workdir[workdirvar])
projectname = "P2Y6KO_WT1"

#Raw Seurat Pipeline
sample.data <- Read10X(data.dir = paste(workdir[workdirvar],
"/filtered_feature_bc_matrix/", sep=""))

sample <- CreateSeuratObject(counts = sample.data, min.cells = 0,
min.features = 100, project = projectname, assay = "RNA")
rm(sample.data)

sample[["percent.mt"]] <- PercentageFeatureSet(sample, pattern = "^mt-")
VlnPlot(sample, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"),
ncol = 3, group.by = "orig.ident")
plot1 <- FeatureScatter(sample, feature1 = "nCount_RNA", feature2 =
"percent.mt")
plot2 <- FeatureScatter(sample, feature1 = "nCount_RNA", feature2 =
"nFeature_RNA")
plot1 + plot2
sample <- subset(sample, subset = nFeature_RNA > 100 & nFeature_RNA < 2500 &
nCount_RNA < 10000 & percent.mt < 20)

sample <- NormalizeData(object = sample, normalization.method =
"LogNormalize", scale.factor = 10000)
sample <- FindVariableFeatures(sample, selection.method = "vst", nfeatures =
2000)

```

```

all.genes <- rownames(sample)
sample <- ScaleData(sample, features = all.genes)
sample <- RunPCA(sample, features = VariableFeatures(object = sample))
sample <- FindNeighbors(sample, dims = 1:15)
sample <- FindClusters(sample, resolution = 1.5)
sample <- RunUMAP(sample, reduction = "pca", dims = 1:15)

#Transfer and setting cluster info, used to assist SoupX downstream
clusterinfo = sample@active.ident
umaploc = sample@reductions[["umap"]][@cell.embeddings

sample_soup = load10X(getwd())
sample_soup = setClusters(sample_soup, setNames(clusterinfo,
rownames(clusterinfo))) #assign cluster from metadata
sample_soup = setDR(sample_soup, umaploc[colnames(sample_soup$toc),])

#Visualise the ratio of observed counts for a gene (or set of genes) to this
expectation value
RhoBefore = plotMarkerMap(sample_soup, "Rho")

#Automated estimation of soup: Rho of sample = 0.06
sample_soup = autoEstCont(sample_soup)

#Correcting expression profile
sample_SoupXout = adjustCounts(sample_soup)

#Create Seurat cleaned-up object
sample_cleaned = CreateSeuratObject(sample_SoupXout)
sample_cleaned <- NormalizeData(object = sample_cleaned, normalization.method
= "LogNormalize", scale.factor = 10000)
sample_cleaned <- FindVariableFeatures(sample_cleaned, selection.method =
"vst", nfeatures = 2000)
all.genes <- rownames(sample_cleaned)
sample_cleaned <- ScaleData(sample_cleaned, features = all.genes)
sample_cleaned <- RunPCA(sample_cleaned, features = VariableFeatures(object =
sample_cleaned))
sample_cleaned <- FindNeighbors(sample_cleaned, dims = 1:15)
sample_cleaned <- FindClusters(sample_cleaned, resolution = 1.5)
sample_cleaned <- RunUMAP(sample_cleaned, reduction = "pca", dims = 1:15)

#Plots
DimPlot(sample, reduction = "umap", label = TRUE, pt.size = 0.6)
FeaturePlot(object = sample, features = "Rho", cols = c("yellow", "blue"))
#Rod
plot(RhoBefore)
plot(RhoAfter)
plotChangeMap(sample_soup, sample_SoupXout, "Rho")
DimPlot(sample_cleaned, reduction = "umap", label = TRUE, pt.size = 0.5)
FeaturePlot(object = sample_cleaned, features = "Rho", cols = c("yellow",
"blue")) #Rod

VlnPlot(sample, features = c("nFeature_RNA", "nCount_RNA"), ncol = 2)

```

X.4 R code of P2Y6-/- experiment data integration

```
library(Seurat)
library(ggplot2)
library(SingleR)
library(dplyr)
library(cellidex)
library(RColorBrewer)
library(SingleCellExperiment)
library(SoupX)
memory.limit(size = 10e12)

#Directories of samples
workdir = c("C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_WT1", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_WT2", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_WT3", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_KO1", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_KO2", "C:/Alex/UCL-
Ophthalmology/SingleCellWork/P2Y6KO/P2Y6_KO3")

#Extracting SoupX output
setwd(workdir[1])
load("SoupXpipelineOut.Rdata")
WT1 = sample_cleaned
WT1$orig.ident = "P2Y6_WT1"

setwd(workdir[2])
load("SoupXpipelineOut.Rdata")
WT2 = sample_cleaned
WT2$orig.ident = "P2Y6_WT2"

setwd(workdir[3])
load("SoupXpipelineOut.Rdata")
WT3 = sample_cleaned
WT3$orig.ident = "P2Y6_WT3"

setwd(workdir[4])
load("SoupXpipelineOut.Rdata")
KO1 = sample_cleaned
KO1$orig.ident = "P2Y6_KO1"

setwd(workdir[5])
load("SoupXpipelineOut.Rdata")
KO2 = sample_cleaned
KO2$orig.ident = "P2Y6_KO2"

setwd(workdir[6])
load("SoupXpipelineOut.Rdata")
KO3 = sample_cleaned
KO3$orig.ident = "P2Y6_KO3"

rm(RhoAfter, RhoBefore, sample, sample_cleaned, sample_soup)

#Normalizing each samples
ALL.list <- c(WT1, WT2, WT3, KO1, KO2, KO3)
```

```

for (i in 1:length(ALL.list)) {
  ALL.list[[i]] <- NormalizeData(ALL.list[[i]], verbose = FALSE)
  ALL.list[[i]] <- FindVariableFeatures(ALL.list[[i]], selection.method =
"vst",
                                     nfeatures = 2000, verbose =
FALSE)
}

#Integration
ALL.anchors <- FindIntegrationAnchors(object.list = ALL.list, dims = 1:50,
reduction = "rpca" )
ALL.integrated <- IntegrateData(anchorset = ALL.anchors, dims = 1:50)
DefaultAssay(ALL.integrated) <- "integrated"

#Scaling
ALL.integrated <- ScaleData(ALL.integrated, verbose = FALSE)

#PCA
ALL.integrated <- RunPCA(ALL.integrated, npcs = 50, verbose = TRUE)
DimPlot(object = ALL.integrated, reduction = "pca")
ElbowPlot(object = ALL.integrated, ndims = 50)

#UMAP
ALL.integrated <- RunUMAP(ALL.integrated, reduction = "pca", dims = 1:25)

ALL.overlapPlot <- DimPlot(ALL.integrated, reduction = "umap", group.by =
"orig.ident")
ALL.overlapPlot

#Clustering
ALL.integrated <- FindNeighbors(ALL.integrated, dims = 1:25)
#a resolution parameter that sets the 'granularity' of the downstream
clustering, with increased values leading to a greater number of clusters. We
find that setting this parameter between 0.4-1.2 typically returns good
results for single cell datasets of around 3K cells.
ALL.integrated <- FindClusters(ALL.integrated, resolution = 1.1)

#Plots
DimPlot(ALL.integrated, reduction = "umap", label = FALSE, pt.size = 0.5,
group.by = "orig.ident", raster = FALSE)
DimPlot(ALL.integrated, reduction = "umap", label = TRUE, pt.size = 0.5,
raster = FALSE)
plot1 <- DimPlot(ALL.integrated, reduction = "umap", label = FALSE, pt.size =
0.6, raster=FALSE)
LabelClusters(plot1, id = "ident", color = "black", size = 6, repel = T,
box.padding = 0.2)

ALL.integrated.markers <- FindAllMarkers(ALL.integrated, only.pos = FALSE,
min.pct = 0, logfc.threshold = 0)
write.csv(ALL.integrated.markers, "P2Y6_ALLclustermarker.csv")

#Cluster renaming after identification
new.cluster.ids <- c("0-Mito Contam", "1-ROD", "2-ROD", "3-ROD", "4-ROD", "5-
MÜLLER GLIA", "6-BIPOLAR", "7-ROD", "8-ROD", "9-ROD", "10-MÜLLER GLIA", "11-

```

```

RGC", "12-ROD", "13-CONE", "14-ROD", "15-ROD", "16-ROD", "17-ROD", "18-
RGC/AMACRINE/BIPOLAR", "19-ROD", "20-ROD", "21-replicating cells", "22-
replicating cells", "23-IMMUNE CELLS", "24-amacrine/bipolar", "25-BIPOLAR", "26-
PRE PHOTORECEPTORS", "27-CONE", "28-Endothelium", "29-RPE")
names(new.cluster.ids) <- levels(ALL.integrated)
ALL.integrated.rename <- RenameIds(ALL.integrated, new.cluster.ids)

plot1 <- DimPlot(ALL.integrated.rename, reduction = "umap", label = FALSE,
pt.size = 0.6, raster=FALSE)

LabelClusters(plot1, id = "ident", color = "black", size = 5, repel = T,
box.padding = 0.8)

#RGC Subset: 11 + 18 -----
ALL.integrated.mixedgroup = subset(ALL.integrated, ids = c(11,18))

DimPlot(ALL.integrated.mixedgroup, reduction = "umap", label = FALSE, pt.size
= 0.6, raster=FALSE)

#MG Subset: 5+10 -----
VlnPlot(object = ALL.integrated, features =
c("rna_Glul", "rna_Dbi", "rna_Clu", "rna_Apoe")) #RGC, Opn4, Pou4f1

ALL.integrated.MG = subset(ALL.integrated, ids = c(5,10))
ALL.integrated.MG[["old.ident"]] <- Idents(object = ALL.integrated.MG)
DimPlot(ALL.integrated.MG, reduction = "umap", label = FALSE, pt.size =
0.6, raster=FALSE)

#Differential expression
WT.integrated = subset(ALL.integrated, orig.ident == "P2Y6_WT1"|orig.ident
=="P2Y6_WT2"|orig.ident == "P2Y6_WT3")
DimPlot(WT.integrated, reduction = "umap", label = TRUE, pt.size = 0.5)
Idents(WT.integrated) = "WT"

KO.integrated = subset(ALL.integrated, orig.ident == "P2Y6_KO1"|orig.ident
=="P2Y6_KO2"|orig.ident == "P2Y6_KO3")
DimPlot(KO.integrated, reduction = "umap", label = TRUE, pt.size = 0.5)
Idents(KO.integrated) = "KO"

ALL.combined <- merge(x= WT.integrated, y = KO.integrated)

ALL.combined.markers2 <- FindMarkers(ALL.combined, ident.1 = "KO", ident.2 =
"WT", only.pos = FALSE, min.pct = 0, logfc.threshold = 0)
write.csv(ALL.combined.markers2, "P2Y6_ALL_DiffExp2.csv")

```

Bibliography

- Alles, J., Karaiskos, N., Praktijnjo, S. D., Grosswendt, S., Wahle, P., Ruffault, P. L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., Zinzen, R., Kocks, C., & Rajewsky, N. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biology*, 15(1), 1–14. <https://doi.org/10.1186/S12915-017-0383-5/FIGURES/4>
- Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, 8(9), 1765–1786. <https://doi.org/10.1038/NPROT.2013.099>
- Andrews, T. S., & Hemberg, M. (2018). Identifying cell populations with scRNASeq. *Molecular Aspects of Medicine*, 59, 114–122. <https://doi.org/10.1016/J.MAM.2017.07.002>
- Awadalla, M. S., Fingert, J. H., Roos, B. E., Chen, S., Holmes, R., Graham, S. L., Chehade, M., Galanopolous, A., Ridge, B., Souzeau, E., Zhou, T., Siggs, O. M., Hewitt, A. W., Mackey, D. A., Burdon, K. P., & Craig, J. E. (2015). Copy Number Variations of TBK1 in Australian Patients With Primary Open-Angle Glaucoma. *American Journal of Ophthalmology*, 159(1), 124-130.e1. <https://doi.org/10.1016/J.AJO.2014.09.044>
- B A Bejjani, D. W. (2000). Multiple CYP1B1 mutations and incomplete penetrance in an inbred population segregating primary congenital glaucoma suggest frequent de novo events and a dominant modifier locus.
- Bacher, R., Chu, L. F., Leng, N., Gasch, A. P., Thomson, J. A., Stewart, R. M., Newton, M., & Kendzierski, C. (2017). SCnorm: robust normalization of single-cell RNA-seq data. *Nature Methods* 2017 14:6, 14(6), 584–586. <https://doi.org/10.1038/nmeth.4263>
- Bacher, R., & Kendzierski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biology*, 17(1), 1–14. <https://doi.org/10.1186/S13059-016-0927-Y/TABLES/1>
- Bajenaru, M. L., Zhu, Y., Hedrick, N. M., Donahoe, J., Parada, L. F., & Gutmann, D. H. (2002). Astrocyte-specific inactivation of the neurofibromatosis 1 gene (NF1) is insufficient for astrocytoma formation. *Molecular and Cellular Biology*, 22(14), 5100–5113. <https://doi.org/10.1128/MCB.22.14.5100-5113.2002>
- Bar, I., Guns, P. J., Metallo, J., Cammarata, D., Wilkin, F., Boeynants, J. M., Bult, H., & Robaye, B. (2008). Knockout mice reveal a role for P2Y6 receptor in macrophages, endothelial cells, and vascular smooth muscle cells. *Molecular Pharmacology*, 74(3), 777–784. <https://doi.org/10.1124/MOL.108.046904>
- Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 2018 37:1, 37(1), 38–44. <https://doi.org/10.1038/nbt.4314>
- Birch, J., & Gil, J. (2020). *Senescence and the SASP: many therapeutic avenues*. <https://doi.org/10.1101/gad.343129>
- Bonham, L. W., Karch, C. M., Fan, C. C., Tan, C., Geier, E. G., Wang, Y., Wen, N., Broce, I. J., Li, Y., Barkovich, M. J., Ferrari, R., Hardy, J., Momeni, P., Höglinger, G., Müller, U., Hess, C. P., Sugrue, L.

- P., Dillon, W. P., Schellenberg, G. D., ... Singleton, A. B. (2018). CXCR4 involvement in neurodegenerative diseases. *Translational Psychiatry* 2017 8:1, 8(1), 1–10. <https://doi.org/10.1038/s41398-017-0049-7>
- Brennecke, P., Anders, S., Kim, J. K., Kołodziejczyk, A. A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S. A., Marioni, J. C., & Heisler, M. G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nature Methods* 2013 10:11, 10(11), 1093–1095. <https://doi.org/10.1038/nmeth.2645>
- Bringmann, A., & Reichenbach, A. (2001). Role of Muller cells in retinal degenerations. *Frontiers in Bioscience : A Journal and Virtual Library*, 6. <https://doi.org/10.2741/BRINGMAN>
- Burgoyne, C. F., Crawford Downs, J., Bellezza, A. J., Francis Suh, J. K., & Hart, R. T. (2005). The optic nerve head as a biomechanical structure: A new paradigm for understanding the role of IOP-related stress and strain in the pathophysiology of glaucomatous optic nerve head damage. In *Progress in Retinal and Eye Research* (Vol. 24, Issue 1, pp. 39–73). Prog Retin Eye Res. <https://doi.org/10.1016/j.preteyeres.2004.06.001>
- Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A., & Theis, F. J. (2018). A test metric for assessing single-cell RNA-seq batch correction. *Nature Methods* 2018 16:1, 16(1), 43–49. <https://doi.org/10.1038/s41592-018-0254-1>
- Casneuf, T., van de Peer, Y., & Huber, W. (2007). In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation. *BMC Bioinformatics*, 8. <https://doi.org/10.1186/1471-2105-8-461>
- Casson, R. J., Chidlow, G., Wood, J. P. M., Crowston, J. G., & Goldberg, I. (2012). Definition of glaucoma: clinical and experimental concepts. *Clinical & Experimental Ophthalmology*, 40(4), 341–349. <https://doi.org/10.1111/J.1442-9071.2012.02773.X>
- Castellano, B. M., Thelen, A. M., Moldavski, O., Feltes, M., van der Welle, R. E. N., Mydock-McGrane, L., Jiang, X., van Eijkeren, R. J., Davis, O. B., Louie, S. M., Perera, R. M., Covey, D. F., Nomura, D. K., Ory, D. S., & Zoncu, R. (2017). Lysosomal cholesterol activates mTORC1 via an SLC38A9-Niemann-Pick C1 signaling complex. *Science (New York, N.Y.)*, 355(6331), 1306–1311. <https://doi.org/10.1126/SCIENCE.AAG1417>
- Chai, A. B., Ammit, A. J., & Gelissen, I. C. (2017). Examining the role of ABC lipid transporters in pulmonary lipid homeostasis and inflammation. *Respiratory Research*, 18(1). <https://doi.org/10.1186/S12931-017-0526-9>
- Chen, Y., Lin, Y., Vithana, E. N., Jia, L., Zuo, X., Wong, T. Y., Chen, L. J., Zhu, X., Tam, P. O. S., Gong, B., Qian, S., Li, Z., Liu, X., Mani, B., Luo, Q., Guzman, C., Leung, C. K. S., Li, X., Cao, W., ... Yang, Z. (2014). Common variants near ABCA1 and in PMM2 are associated with primary open-angle glaucoma. *Nature Genetics*, 46(10), 1115–1119. <https://doi.org/10.1038/ng.3078>
- Chunchun Hu, L. N. (2020). ABCA1 Regulates IOP by Modulating Cav1/eNOS/NO Signaling Pathway. *Invest Ophthalmol Vis Sci*.

- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature* 1970 227:5258, 227(5258), 561–563.
<https://doi.org/10.1038/227561a0>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. <https://doi.org/10.1093/BIOINFORMATICS/BTS635>
- Eberle, D., Santos-Ferreira, T., Grahl, S., & Ader, M. (2014). Subretinal Transplantation of MACS Purified Photoreceptor Precursor Cells into the Adult Mouse Retina. *Journal of Visualized Experiments : JoVE*, 84(84), 50932. <https://doi.org/10.3791/50932>
- Ester Reina-Torres, J. C.-K. (2017). VEGF as a Paracrine Regulator of Conventional Outflow Facility. *Invest Ophthalmol Vis Sci*, 1899.
- Feodorova, Y., Koch, M., Bultman, S., Michalakakis, S., & Solovei, I. (2015). Quick and reliable method for retina dissociation and separation of rod photoreceptor perikarya from adult mice. *MethodsX*, 2, 39–46. <https://doi.org/10.1016/J.MEX.2015.01.002>
- Fletcher, E. L., Downie, L. E., Ly, A., Ward, M. M., Batcha, A. H., Puthussery, T., Yee, P., & Hatzopoulos, K. M. (2021). A review of the role of glial cells in understanding retinal disease.
<https://doi.org/10.1111/j.1444-0938.2007.00204.x>, 91(1), 67–77. <https://doi.org/10.1111/J.1444-0938.2007.00204.X>
- Friedlander, M. (2007). Fibrosis and diseases of the eye. *The Journal of Clinical Investigation*, 117(3), 576–586. <https://doi.org/10.1172/JCI31030>
- Fujimoto, T., Inoue, T., Maki, K., Kojima, S., Inoue-Mochita, M., & Tanihara, H. (2015). The Effect of VEGF-A on Trabecular Meshwork and Schlemm's Canal Endothelial Cells Permeability. *Investigative Ophthalmology & Visual Science*.
- Fujino, Y., Asaoka, R., Murata, H., Miki, A., Tanito, M., Mizoue, S., Mori, K., Suzuki, K., Yamashita, T., Kashiwagi, K., Shoji, N., Matsuura, M., Yanagisawa, M., Hirasawa, H., Mayama, C., Usui, S., Matsushita, K., Nishida, K., Omura, T., ... Hirasawa, K. (2016). Evaluation of Glaucoma Progression in Large-Scale Clinical Data: The Japanese Archive of Multicenter Databases in Glaucoma (JAMDIG). *Investigative Ophthalmology & Visual Science*, 57(4), 2012–2020.
<https://doi.org/10.1167/IOVS.15-19046>
- Geisert, E. E., Lu, L., Freeman-Anderson, N. E., Templeton, J. P., Nassr, M., Wang, X., Gu, W., Jiao, Y., & Williams, R. W. (2009). Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. *Molecular Vision*, 15, 1730. <https://pmc/articles/PMC2736153/>
- Gharahkhani, P., Burdon, K. P., Fogarty, R., Sharma, S., Hewitt, A. W., Martin, S., Law, M. H., Cremin, K., Bailey, J. N. C., Loomis, S. J., Pasquale, L. R., Haines, J. L., Hauser, M. A., Viswanathan, A. C., McGuffin, P., Topouzis, F., Foster, P. J., Graham, S. L., Casson, R. J., ... Craig, J. E. (2014). Common variants near ABCA1, AFAP1 and GMD5 confer risk of primary open-angle glaucoma. *Nature Genetics*, 46(10), 1120–1125. <https://doi.org/10.1038/ng.3079>
- Gharahkhani, P., Jorgenson, E., Hysi, P., Khawaja, A. P., Pendergrass, S., Han, X., Ong, J. S., Hewitt, A. W., Segre, A. v., Rouhana, J. M., Hamel, A. R., Igo, R. P., Choquet, H., Qassim, A., Josyula, N. S., Cooke

- Bailey, J. N., Bonnemaier, P. W. M., Iglesias, A., Siggs, O. M., ... Wiggs, J. L. (2021). Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nature Communications* 2021 12:1, 12(1), 1–16. <https://doi.org/10.1038/s41467-020-20851-4>
- Glial cells of the Retina by Helga Kolb – Webvision.* (n.d.). Retrieved October 27, 2020, from <https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/glial-cells-of-the-retina/>
- Goel, M. (2010). Aqueous Humor Dynamics: A Review~!2010-03-03~!2010-06-17~!2010-09-02~! *The Open Ophthalmology Journal*, 4(1), 52–59. <https://doi.org/10.2174/1874364101004010052>
- Grotemeyer, A., McFleder, R. L., Wu, J., Wischhusen, J., & Ip, C. W. (2022). Neuroinflammation in Parkinson's Disease - Putative Pathomechanisms and Targets for Disease-Modification. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/FIMMU.2022.878771>
- Haghverdi, L., Lun, A. T. L., Morgan, M. D., & Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* 2018 36:5, 36(5), 421–427. <https://doi.org/10.1038/nbt.4091>
- Hamon, Y., Broccardo, C., Chambenoit, O., Luciani, M.-F., Toti, F., Chaslin, S., Freyssinet, J.-M., Devaux, P. F., Mcneish, J., Marguet, D., & Chimini, G. (2000). ABC1 promotes engulfment of apoptotic cells and transbilayer redistribution of phosphatidylserine. *NATURE CELL BIOLOGY*, 2, 399. www.nature.com/ncb
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13), 3573-3587.e29. <https://doi.org/10.1016/J.CELL.2021.04.048>
- Heneka, M. T., Carson, M. J., Khoury, J. el, Landreth, G. E., Brosseon, F., Feinstein, D. L., Jacobs, A. H., Wyss-Coray, T., Vitorica, J., Ransohoff, R. M., Herrup, K., Frautschy, S. A., Finsen, B., Brown, G. C., Verkhratsky, A., Yamanaka, K., Koistinaho, J., Latz, E., Halle, A., ... Kummer, M. P. (2015). Neuroinflammation in Alzheimer's Disease. *The Lancet. Neurology*, 14(4), 388. [https://doi.org/10.1016/S1474-4422\(15\)70016-5](https://doi.org/10.1016/S1474-4422(15)70016-5)
- Hennis, A., Wu, S. Y., Nemesure, B., Honkanen, R., & Leske, M. C. (2007). Awareness of incident open-angle glaucoma in a population study: the Barbados Eye Studies. *Ophthalmology*, 114(10), 1816–1821. <https://doi.org/10.1016/J.OPHTHA.2007.06.013>
- Hu, J., Popp, R., Frömel, T., Ehling, M., Awwad, K., Adams, R. H., Hammes, H. P., & Fleming, I. (2014). Müller glia cells regulate Notch signaling and retinal angiogenesis via the generation of 19,20-dihydroxydocosapentaenoic acid. *Journal of Experimental Medicine*, 211(2), 281–295. <https://doi.org/10.1084/JEM.20131494>
- Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, 50(8). <https://doi.org/10.1038/S12276-018-0071-8>
- Hysi, P. G., Cheng, C. Y., Springelkamp, H., Macgregor, S., Cooke Bailey, J. N., Wojciechowski, R., Vitart, V., Nag, A., Hewitt, A. W., Höhn, R., Venturini, C., Mirshahi, A., Ramdas, W. D., Thorleifsson, G.,

- Vithana, E., Khor, C. C., Stefansson, A. B., Liao, J., Haines, J. L., ... Vukcevic, D. (2014). Genome-wide analysis of multi-ancestry cohorts identifies new loci influencing intraocular pressure and susceptibility to glaucoma. *Nature Genetics*, 46(10), 1126–1130. <https://doi.org/10.1038/ng.3087>
- Iwase, A., Suzuki, Y., Araie, M., Yamamoto, T., Abe, H., Shirato, S., Kuwayama, Y., Mishima, H. K., Shimizu, H., Tomita, G., Inoue, Y., & Kitazawa, Y. (2004). The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. *Ophthalmology*, 111(9), 1641–1648. <https://doi.org/10.1016/J.OPHTHA.2004.03.029>
- Jiao, H., Rutar, M., Fernando, N., Yednock, T., Sankaranarayanan, S., Aggio-Bruce, R., Provis, J., & Natoli, R. (2018). Subretinal macrophages produce classical complement activator C1q leading to the progression of focal retinal degeneration. *Molecular Neurodegeneration*, 13(1), 45. <https://doi.org/10.1186/s13024-018-0278-0>
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., & Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 36(Database issue). <https://doi.org/10.1093/NAR/GKM882>
- Kang, J. H., Boumenna, T., Stein, J. D., Khawaja, A., Rosner, B. A., Wiggs, J. L., & Pasquale, L. R. (2019). Association of Statin Use and High Serum Cholesterol Levels With Risk of Primary Open-Angle Glaucoma. *JAMA Ophthalmology*, 137(7), 756–765. <https://doi.org/10.1001/JAMAOPHTHALMOL.2019.0900>
- Kapetanakis, V. v., Chan, M. P. Y., Foster, P. J., Cook, D. G., Owen, C. G., & Rudnicka, A. R. (2016). Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): a systematic review and meta-analysis. *The British Journal of Ophthalmology*, 100(1), 86–93. <https://doi.org/10.1136/BJOPHTHALMOL-2015-307223>
- Karasinska, J. M., de Haan, W., Franciosi, S., Ruddie, P., Fan, J., Kruit, J. K., Stukas, S., Lütjohann, D., Gutmann, D. H., Wellington, C. L., & Hayden, M. R. (2013). ABCA1 influences neuroinflammation and neuronal death. *Neurobiology of Disease*, 54, 445–455. <https://doi.org/10.1016/J.NBD.2013.01.018>
- Kass, M. A., Heuer, D. K., Higginbotham, E. J., Johnson, C. A., Keltner, J. L., Philip Miller, J., Parrish, R. K., Roy Wilson, M., & Gordon, M. O. (2002). The Ocular Hypertension Treatment Study: A Randomized Trial Determines That Topical Ocular Hypotensive Medication Delays or Prevents the Onset of Primary Open-Angle Glaucoma. *Archives of Ophthalmology*, 120(6), 701–713. <https://doi.org/10.1001/ARCHOPHT.120.6.701>
- Katayama, S., Töhönen, V., Linnarsson, S., & Kere, J. (2013). SAMstrt: statistical test for differential expression in single-cell transcriptome with spike-in normalization. *Bioinformatics*, 29(22), 2943–2945. <https://doi.org/10.1093/BIOINFORMATICS/BTT511>
- Kharchenko, P. v., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods* 2014 11:7, 11(7), 740–742. <https://doi.org/10.1038/nmeth.2967>

- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 2019 37:8, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R., & Hemberg, M. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nature Methods* 2017 14:5, 14(5), 483–486. <https://doi.org/10.1038/nmeth.4236>
- Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., & Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.)*, 308(5720), 385–389. <https://doi.org/10.1126/SCIENCE.1109557>
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 2015(11), 951. <https://doi.org/10.1101/PDB.TOP084970>
- L Lun, A. T., Riesenfeld, S., Andrews, T., Phuong Dao, T., Gomes, T., in the, participants, Human Cell Atlas Jamboree, st, & Marioni, J. C. (n.d.). *EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data*. <https://doi.org/10.1186/s13059-019-1662-y>
- Lafzi, A., Moutinho, C., Picelli, S., & Heyn, H. (2018). Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nature Protocols* 2018 13:12, 13(12), 2742–2757. <https://doi.org/10.1038/s41596-018-0073-y>
- Langmann, T. (2007). Microglia activation in retinal degeneration. *Journal of Leukocyte Biology*, 81(6), 1345–1351. <https://doi.org/10.1189/jlb.0207114>
- Leite, M. T., Sakata, L. M., & Medeiros, F. A. (2011). Managing glaucoma in developing countries. *Arquivos Brasileiros de Oftalmologia*, 74(2), 83–84. <https://doi.org/10.1590/S0004-27492011000200001>
- Lewandowski, C. T., Laham, M. S., & Thatcher, G. R. J. (2022). Remembering your A, B, C's: Alzheimer's disease and ABCA1. *Acta Pharmaceutica Sinica B*, 12(3), 995–1018. <https://doi.org/10.1016/J.APSB.2022.01.011>
- Li, A., Zhang, X., Zheng, D., Ge, J., Laties, A. M., & Mitchell, C. H. (2011). Sustained elevation of extracellular ATP in aqueous humor from humans with primary chronic angle-closure glaucoma. *Experimental Eye Research*, 93(4), 528–533. <https://doi.org/10.1016/J.EXER.2011.06.020>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7), 923–930. <https://doi.org/10.1093/BIOINFORMATICS/BTT656>
- Limbad, C., Oron, T. R., Alimirah, F., Davalos, A. R., Tracy, T. E., Gan, L., Desprez, P. Y., & Campisi, J. (2020). Astrocyte senescence promotes glutamate toxicity in cortical neurons. *PLOS ONE*, 15(1), e0227887. <https://doi.org/10.1371/JOURNAL.PONE.0227887>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9>

- Lukowski, S. W., Lo, C. Y., Sharov, A. A., Nguyen, Q., Fang, L., Hung, S. S., Zhu, L., Zhang, T., Grünert, U., Nguyen, T., Senabouth, A., Jabbari, J. S., Welby, E., Sowden, J. C., Waugh, H. S., Mackey, A., Pollock, G., Lamb, T. D., Wang, P., ... Wong, R. C. (2019). A single-cell transcriptome atlas of the adult human retina. *The EMBO Journal*, 38(18). <https://doi.org/10.15252/embj.2018100811>
- Lupton, M. K., Proitsi, P., Lin, K., Hamilton, G., Daniilidou, M., Tsolaki, M., & Powell, J. F. (2014). The role of ABCA1 gene sequence variants on risk of Alzheimer's disease. *Journal of Alzheimer's Disease : JAD*, 38(4), 897–906. <https://doi.org/10.3233/JAD-131121>
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015a). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5), 1202. <https://doi.org/10.1016/J.CELL.2015.05.002>
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015b). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5), 1202–1214. <https://doi.org/10.1016/J.CELL.2015.05.002>
- Madeira, M. H., Boia, R., Santos, P. F., Ambrósio, A. F., & Santiago, A. R. (2015). Contribution of microglia-mediated neuroinflammation to retinal degenerative diseases. In *Mediators of Inflammation* (Vol. 2015). Hindawi Limited. <https://doi.org/10.1155/2015/673090>
- Mallick, J., Devi, L., Malik, P. K., & Mallick, J. (2016). Update on Normal Tension Glaucoma. *Journal of Ophthalmic & Vision Research*, 11(2), 204–208. <https://doi.org/10.4103/2008-322X.183914>
- Manir Ali, M. M.-L. (2009). Null mutations in LTBP2 cause primary congenital glaucoma.
- Markovskaya, A., Crooke, A., Guzmán-Aranguéz, A. I., Peral, A., Ziganshin, A. U., & Pintor, J. (2008). Hypotensive effect of UDP on intraocular pressure in rabbits. *European Journal of Pharmacology*, 579(1–3), 93–97. <https://doi.org/10.1016/J.EJPHAR.2007.10.040>
- Martinez, C. C. (2022). The Interplay of Endothelial P2Y Receptors in Cardiovascular Health: From Vascular Physiology to Pathology. *International Journal of Molecular Sciences*.
- Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience* 2001 4:9, 4(9), 877–886. <https://doi.org/10.1038/nn0901-877>
- Masuda, T., Amann, L., Sankowski, R., Staszewski, O., Lenz, M., d'Errico, P., Snaidero, N., Costa Jordão, M. J., Böttcher, C., Kierdorf, K., Jung, S., Priller, J., Misgeld, T., Vlachos, A., Luehmann, M. M., Knobloch, K. P., & Prinz, M. (2020). Novel Hexb-based tools for studying microglia in the CNS. *Nature Immunology*, 21(7), 802–815. <https://doi.org/10.1038/s41590-020-0707-4>
- McGwin, G., McNeal, S., Owsley, C., Girkin, C., Epstein, D., & Lee, P. P. (2004). Statins and other cholesterol-lowering medications and the presence of glaucoma. *Archives of Ophthalmology (Chicago, Ill. : 1960)*, 122(6), 822–826. <https://doi.org/10.1001/ARCHOPHT.122.6.822>
- Menon, M., Mohammadi, S., Davila-Velderrain, J., Goods, B. A., Cadwell, T. D., Xing, Y., Stemmer-Rachamimov, A., Shalek, A. K., Love, J. C., Kellis, M., & Hafler, B. P. (2019). Single-cell transcriptomic

- atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nature Communications* 2019 10:1, 10(1), 1–9. <https://doi.org/10.1038/s41467-019-12780-8>
- Oram, J. F. (2000). Tangier disease and ABCA1. *Biochimica et Biophysica Acta*, 1529(1–3), 321–330. [https://doi.org/10.1016/S1388-1981\(00\)00157-8](https://doi.org/10.1016/S1388-1981(00)00157-8)
- Phan, H. van, van Gent, M., Drayman, N., Basu, A., Gack, M. U., & Tay, S. (2021). High-throughput RNA sequencing of paraformaldehyde-fixed single cells. *Nature Communications* 2021 12:1, 12(1), 1–11. <https://doi.org/10.1038/s41467-021-25871-2>
- Pintor, J., Sánchez-Nogueiro, J., Irazu, M., Mediero, A., Peláez, T., & Peral, A. (2004). Immunolocalisation of P2Y receptors in the rat eye. *Purinergic Signalling*, 1(1), 83–90. <https://doi.org/10.1007/S11302-004-5072-5>
- Pirro G Hysi, C.-Y. C. (2014). Genome-wide analysis of multi-ancestry cohorts identifies new loci influencing intraocular pressure and susceptibility to glaucoma. *Nature Genetics*, 46.
- Posch-Pertl, L., Michelitsch, M., Wagner, G., Wildner, B., Silbernagel, G., Pregartner, G., & Wedrich, A. (2022). Cholesterol and glaucoma: a systematic review and meta-analysis. *Acta Ophthalmologica*, 100(2), 148–158. <https://doi.org/10.1111/AOS.14769>
- Primary Open-Angle Glaucoma* - EyeWiki. (n.d.). Retrieved October 31, 2022, from https://eyewiki.aao.org/Primary_Open-Angle_Glaucoma
- Quigley, H., & Broman, A. T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. In *British Journal of Ophthalmology* (Vol. 90, Issue 3, pp. 262–267). Br J Ophthalmol. <https://doi.org/10.1136/bjo.2005.081224>
- Rashid, K., Akhtar-Schaefer, I., & Langmann, T. (2019). Microglia in retinal degeneration. In *Frontiers in Immunology* (Vol. 10, Issue AUG). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2019.01975>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. <https://doi.org/10.1093/BIOINFORMATICS/BTP616>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), 1–9. <https://doi.org/10.1186/GB-2010-11-3-R25/FIGURES/3>
- Roy, S., Amin, S., & Roy, S. (2016). Retinal Fibrosis in Diabetic Retinopathy. *Experimental Eye Research*, 142, 71. <https://doi.org/10.1016/J.EXER.2015.04.004>
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* 2015 33:5, 33(5), 495–502. <https://doi.org/10.1038/nbt.3192>
- Schmidt, S. I., Blaabjerg, M., Freude, K., & Meyer, M. (2022). RhoA Signaling in Neurodegenerative Diseases. *Cells*, 11(9). <https://doi.org/10.3390/CELLS11091520>

- Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T., & Pfeiffer, N. (2020). The Diagnosis and Treatment of Glaucoma. *Deutsches Ärzteblatt International*, 117(13), 225. <https://doi.org/10.3238/ARZTEBL.2020.0225>
- Sene, A., Khan, A. A., Cox, D., Nakamura, R. E. I., Santeford, A., Kim, B. M., Sidhu, R., Onken, M. D., Harbour, J. W., Hagbi-Levi, S., Chowers, I., Edwards, P. A., Baldan, A., Parks, J. S., Ory, D. S., & Apte, R. S. (2013). Impaired cholesterol efflux in senescent macrophages promotes age-related macular degeneration. *Cell Metabolism*, 17(4), 549. <https://doi.org/10.1016/J.CMET.2013.03.009>
- Sharif M. Rumjahn, K. A. (2007). P2y Receptor-Mediated Angiogenesis via Vascular Endothelial Growth Factor Receptor 2 Signaling. *Proc West Pharmacol Soc.* .
- Shekhar, K., Lapan, S. W., Whitney, I. E., Tran, N. M., Macosko, E. Z., Kowalczyk, M., Adiconis, X., Levin, J. Z., Nemesh, J., Goldman, M., McCarroll, S. A., Cepko, C. L., Regev, A., & Sanes, J. R. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell*, 166(5), 1308-1323.e30. <https://doi.org/10.1016/j.cell.2016.07.054>
- Shendure, J. (2008). The beginning of the end for microarrays? *Nature Methods*, 5(7), 585–587. <https://doi.org/10.1038/NMETH0708-585>
- Shinozaki, Y., Kashiwagi, K., Namekata, K., Takeda, A., Ohno, N., Robaye, B., Harada, T., Iwata, T., & Koizumi, S. (2017). Purinergic dysregulation causes hypertensive glaucoma-like optic neuropathy. *JCI Insight*, 2(19). <https://doi.org/10.1172/JCI.INSIGHT.93456>
- Silverman, S. M., & Wong, W. T. (2018). Microglia in the Retina: Roles in Development, Maturity, and Disease. *Annual Review of Vision Science*, 4(1), 45–77. <https://doi.org/10.1146/annurev-vision-091517-034425>
- Soumian, S., Albrecht, C., Davies, A. H., & Gibbs, R. G. J. (2005). ABCA1 and atherosclerosis. *Vascular Medicine (London, England)*, 10(2), 109–119. <https://doi.org/10.1191/1358863X05VM593RA>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019a). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888-1902.e21. <https://doi.org/10.1016/J.CELL.2019.05.031>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019b). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888-1902.e21. <https://doi.org/10.1016/J.CELL.2019.05.031/ATTACHMENT/2F8B9EBE-54E6-43EB-9EF2-949B6BDA8BA2/MMC3.PDF>
- Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (n.d.). *Exponential scaling of single-cell RNA-seq in the last decade*. Retrieved September 13, 2022, from <https://www.neb.com/faqs/2012/11/19/what-is-the-starting-material-i-need-to-use-when->
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* 2019 20:8, 20(8), 467–484. <https://doi.org/10.1038/s41576-019-0127-1>

- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5), 377–382. <https://doi.org/10.1038/nmeth.1315>
- Tang, X., Huang, Y., Lei, J., Luo, H., & Zhu, X. (2019). The single-cell sequencing: New developments and medical applications. *Cell and Bioscience*, 9(1), 1–9. <https://doi.org/10.1186/S13578-019-0314-Y/FIGURES/2>
- Tham, Y. C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., & Cheng, C. Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11), 2081–2090. <https://doi.org/10.1016/J.OPHTHA.2014.05.013>
- Timmins, J. M., Lee, J.-Y., Boudyguina, E., Kluckman, K. D., Brunham, L. R., Mulya, A., Gebre, A. K., Coutinho, J. M., Colvin, P. L., Smith, T. L., Hayden, M. R., Maeda, N., & Parks, J. S. (2005). Targeted inactivation of hepatic Abca1 causes profound hypoalphalipoproteinemia and kidney hypercatabolism of apoA-I. *The Journal of Clinical Investigation*, 115(5), 1333–1342. <https://doi.org/10.1172/JCI23915>
- Tran, N. M., Shekhar, K., Whitney, I. E., Jacobi, A., Benhar, I., Hong, G., Yan, W., Adiconis, X., Arnold, M. E., Lee, J. M., Levin, J. Z., Lin, D., Wang, C., Lieber, C. M., Regev, A., He, Z., & Sanes, J. R. (2019a). Single-Cell Profiles of Retinal Ganglion Cells Differing in Resilience to Injury Reveal Neuroprotective Genes. *Neuron*, 104(6), 1039–1055.e12. <https://doi.org/10.1016/J.NEURON.2019.11.006>
- Tran, N. M., Shekhar, K., Whitney, I. E., Jacobi, A., Benhar, I., Hong, G., Yan, W., Adiconis, X., Arnold, M. E., Lee, J. M., Levin, J. Z., Lin, D., Wang, C., Lieber, C. M., Regev, A., He, Z., & Sanes, J. R. (2019b). Single-Cell Profiles of Retinal Ganglion Cells Differing in Resilience to Injury Reveal Neuroprotective Genes. *Neuron*, 104(6), 1039–1055.e12. <https://doi.org/10.1016/J.NEURON.2019.11.006/ATTACHMENT/E091CDB1-E8A4-45BC-9F1B-517EEC865444/MMC1.PDF>
- Tran, V., Papalexi, E., Schroeder, S., Kim, G., Sapre, A., Pangallo, J., Sova, A., Matulich, P., Kenyon, L., Sayar, Z., Koehler, R., Diaz, D., Gadkari, A., Howitz, K., Nigos, M., Roco, C. M., & Rosenberg, A. B. (n.d.). *High sensitivity single cell RNA sequencing with split pool barcoding*. <https://doi.org/10.1101/2022.08.27.505512>
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31(1), 46–53. <https://doi.org/10.1038/NBT.2450>
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, 25(9), 1105–1111. <https://doi.org/10.1093/BIOINFORMATICS/BTP120>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), 511–515. <https://doi.org/10.1038/NBT.1621>

- van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- van Hove, H., Martens, L., Scheyltjens, I., de Vlaminck, K., Pombo Antunes, A. R., de Prijck, S., Vandamme, N., de Schepper, S., van Isterdael, G., Scott, C. L., Aerts, J., Berx, G., Boeckxstaens, G. E., Vandenbroucke, R. E., Vereecke, L., Moechars, D., Guillems, M., van Ginderachter, J. A., Saeys, Y., & Movahedi, K. (2019). A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment. *Nature Neuroscience*, 22(6), 1021–1035. <https://doi.org/10.1038/s41593-019-0393-4>
- Wang, Q., Liu, Y., & Zhou, J. (2015). Neuroinflammation in Parkinson's disease and its potential as therapeutic target. *Translational Neurodegeneration*, 4(1). <https://doi.org/10.1186/s40035-015-0042-0>
- Wang, R., Yu, Z., Sunchu, B., Shoaf, J., Dang, I., Zhao, S., Caples, K., Bradley, L., Beaver, L. M., Ho, E., Löhr, C. v., & Perez, V. I. (2017). Rapamycin inhibits the secretory phenotype of senescent cells by a Nrf2-independent mechanism. *Aging Cell*, 16(3), 564–574. <https://doi.org/10.1111/ACEL.12587>
- Watanabe, T., & Raff, M. C. (1988). Retinal astrocytes are immigrants from the optic nerve. *Nature*, 332(6167), 834–837. <https://doi.org/10.1038/332834a0>
- Weinreb, R. N., Aung, T., & Medeiros, F. A. (2014). The pathophysiology and treatment of glaucoma: A review. In *JAMA - Journal of the American Medical Association* (Vol. 311, Issue 18, pp. 1901–1911). American Medical Association. <https://doi.org/10.1001/jama.2014.3192>
- Wen, J. C., Reina-Torres, E., Sherwood, J. M., Challa, P., Liu, K. C., Li, G., . . . Allingham, R. R. (2017). Intravitreal Anti-VEGF Injections Reduce Aqueous Outflow Facility in Patients With Neovascular Age-Related Macular Degeneration. *Physiology and Pharmacology*.
- Xiao, Z., & Gong, X. (2017). Use of Statins and Risk of Reducing Glaucoma: Is There a Link? *Medical Principles and Practice*, 26(3), 296. <https://doi.org/10.1159/000453040>
- Xu, S., Cai, Y., & Wei, Y. (2014). mTOR Signaling from Cellular Senescence to Organismal Aging. *Aging and Disease*, 5(4), 263. <https://doi.org/10.14336/AD.2014.0500263>
- Yan, W., Laboulaye, M. A., Tran, N. M., Whitney, I. E., Benhar, I., & Sanes, J. R. (2020). Mouse Retinal Cell Atlas: Molecular Identification of over Sixty Amacrine Cell Types. *Journal of Neuroscience*, 40(27), 5177–5195. <https://doi.org/10.1523/JNEUROSCI.0471-20.2020>
- Young, M., & Behjati, S. (2018). SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. *BioRxiv*, 2020, 303727. <https://doi.org/10.1101/303727>
- Yurco, P., & Cameron, D. A. (2005). Responses of Müller glia to retinal injury in adult zebrafish. *Vision Research*, 45(8), 991–1002. <https://doi.org/10.1016/j.visres.2004.10.022>
- Zhang, M. L., Zhao, G. L., Hou, Y., Zhong, S. M., Xu, L. J., Li, F., Niu, W. R., Yuan, F., Yang, X. L., Wang, Z., & Miao, Y. (2020). Rac1 conditional deletion attenuates retinal ganglion cell apoptosis by accelerating autophagic flux in a mouse model of chronic ocular hypertension. *Cell Death & Disease* 2020 11:9, 11(9), 1–13. <https://doi.org/10.1038/s41419-020-02951-7>

- Zhang, X., Li, A., Ge, J., Reigada, D., Laties, A. M., & Mitchell, C. H. (2007). Acute increase of intraocular pressure releases ATP into the anterior chamber. *Experimental Eye Research*, 85(5), 637–643. <https://doi.org/10.1016/J.EXER.2007.07.016>
- Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O’Keefe, S., Phatnani, H. P., Guarnieri, P., Caneda, C., Ruderisch, N., Deng, S., Liddelow, S. A., Zhang, C., Daneman, R., Maniatis, T., Barres, B. A., & Wu, J. Q. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 34(36), 11929–11947. <https://doi.org/10.1523/JNEUROSCI.1860-14.2014>
- Zhang, Y., & Wallace, B. C. (2015). *A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification*. <https://doi.org/10.48550/arxiv.1510.03820>