Retinal aging transcriptome and cellular landscape in association with the progression of age-related macular degeneration

Jiang-Hui Wang1*, Raymond C.B. Wong1,3, Guei-Sheung Liu1,2,3,4*

1Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia
2Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia
3Ophthalmology, Department of Surgery, University of Melbourne, East Melbourne, Victoria, Australia
4Aier Eye Institute, Changsha, Hunan, China

*Correspondence and requests for materials should be addressed to Dr Jiang-Hui Wang (sloanjhwang@gmail.com). Centre for Eye Research Australia. Address: Level 7, 32 Gisborne Street, East Melbourne, VIC 3002, Australia. Tel: +61399298463.

Or

Dr Guei-Sheung Liu (rickliu0817@gmail.com). Centre for Eye Research Australia. Address: Level 7, 32 Gisborne Street, East Melbourne, VIC 3002, Australia. Tel: +61362264250.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract
Age is the main risk factor for age-related macular degeneration (AMD), a leading cause of blindness in the elderly, with limited therapeutic options. Here we systematically analyzed the transcriptomic characteristics and cellular landscape of the aging retina from controls and patients with AMD. We identify the aging genes in the retina that are associated with innate immune response and inflammation. Deconvolution analysis reveals that the estimated proportion of M2 and M0 macrophages is increased and decreased, respectively with both age and AMD severity. Moreover, we find that Müller glia are increased with age but not with disease severity. Several genes associated with both age and disease severity in AMD, particularly C1s and MR1, are strong positively correlated with the proportions of Müller glia. Our studies expand the genetic and cellular landscape of AMD and provide avenues for further studies on the relationship between age and AMD.
Main

Aging is one of the most important demographic risk factors for age-related macular degeneration (AMD)\textsuperscript{1,2}, the leading causes of blindness among people aged over 55-year-old in the developed world\textsuperscript{3}. A review summarizing the prevalence of AMD over age from various studies suggested a clear trend that the prevalence of AMD regardless mild or severe clinical features is gradually increased with age\textsuperscript{2}. Advances in sequencing technologies, genetic analyses, such as genome-wide association studies (GWAS), using large numbers of samples from donors with AMD, enables identification of common and rare variation that is associated with AMD\textsuperscript{4}. Moreover, expression quantitative trait loci (eQTL) analysis using transcriptional profiles of postmortem retinas from 453 controls and AMD patients expanded the genetic landscape of AMD\textsuperscript{5}. Such studies also have implicated roles for the immune system, in particular the dysregulation of the complement system, in the pathogenesis and severity of AMD\textsuperscript{6}. While these studies provided useful information on possible causal genes or immune cell types associated with AMD, these is little information on how age affects the progression of AMD.

In this study, we systematically assess the transcriptomic profiles, immune and retinal cellular landscapes of the aging retinas in association with the progression of AMD. We define the aging genes in the retina and show that the aging retina is transcriptionally featured by innate immune response and inflammation. Deconvolution analysis reveals increased and decreased proportion of M2 and M0 macrophages, respectively with both age and disease severity during the progression of AMD. We further identify specific genes that altered with age and disease severity, which are relevant to immune response, leukocyte activation and multicellular organismal homeostasis. Our results suggest that Müller glia are increased with age but not with disease severity, further indicating that age is the principal risk factor that uniquely contributes to the severity of AMD. Our study provides pinpointed detailed genetic and cellular characteristics of the aging retina that contribute to the severity of AMD.

Results

We analysed the Eye Genotype Expression (EyeGEx) database, a public resource of retinal transcriptional profiles from patients with AMD\textsuperscript{3}. A total of 453 previously quality-controlled\textsuperscript{2} retinal RNA-seq profiles from controls and cases of different ages (50-109 years old) and at distinct stages of AMD were analyzed (Fig.1a). The Minnesota Grading System (MGS) was used in this database to classify the disease severity. MGS1 categorizes donor
retinas without AMD symptoms and serves as controls, whereas MGS2 to MGS4 categorize donor retinas at the progressively advanced AMD stages.

Identification of age-associated genes in the human retina

Gene expression levels in the EyeGEx database are previously defined by a number of clinical variables, including MGS level, sex and age\(^7\). By controlling for MGS level and sex, therefore, we identified two clusters of genes associated with progressive aging (adjusted \(p<0.05\), Supplementary Data 1) using a likelihood-ratio test\(^8\). Specifically, we found that expression levels of 445 and 57 genes were increased and decreased over age, respectively (Fig. 1b, hereafter referred to Age-up and Age-down genes, respectively). A heatmap of expression profile of all Age-up and Age-down genes also shows the progressive changes of these genes over age (Fig. 1c). Gene ontology (GO biological process) analysis of Age-up genes by Metascape, a comprehensive and combinational knowledgebase\(^9\), indicates that innate immune responses, leukocyte activation and inflammatory responses are among the top GO terms (Fig. 1d, Extended Data Fig. 1, Supplementary Data 2). Given the limited number of Age-down genes, there were no significant GO terms enriched. Our results were consistent with previous reports that retinal aging is a major risk factor of developing various retinal degenerative diseases mainly via dysregulation of innate immune systems, including AMD, diabetic retinopathy and glaucomatous retinopathy\(^10,11\). Studies on the retinal immune cells in particular macrophages suggested its activation in the retina is reported in both AMD patients\(^12\) and old mice\(^13\).

The immune cellular landscape in the retina of patients with AMD

In order to assess different immune cell types and to understand their relationship with age or disease severity (ranked by MGS level), in the retina of patients with AMD, we deconvoluted the EyeGEx database by CIBERSORTx, a machine learning method that enables inference of cell-type-specific gene expression profiles without physical cell isolation\(^14\). LM22, a signature matrix that enables distinction of 22 immune cell types, was used as a reference for deconvolution analysis\(^15\). We specifically looked at how age or MGS level as an individual factor impact the cellular proportions of immune cells using an ordinal logistic regression (OLR) model. By controlling for sex and MGS level, we found that proportions of M1 macrophages, M2 macrophages, monocytes were significantly increased with age, while M0 macrophages were decreased (Extended Data Figs. 2 and 3, Supplementary Data 3). By controlling for sex and age, interestingly, we observed that proportions of activated NK cells
and M2 macrophages were significantly increased with MGS level, while M0 macrophages were also decreased (Extended Data Figs. 4 and 5, Supplementary Data 4). It is known that human peripheral monocytes were differentiated into inactivated M0 macrophages and then polarized to M1 and M2 phenotypes under various stress or stimuli\textsuperscript{16}. Previous studies have shown that numbers of both M1 and M2 macrophages were increased in the retina\textsuperscript{12} or aqueous humors\textsuperscript{17} of a small number of neovascular AMD patients or laser-induced murine model with choroidal neovascularization\textsuperscript{17, 18}. In line with these reports, our results further revealed that age alone as a predominant factor that triggers the polarization of M0 macrophages towards pro-inflammatory M1 and pro-angiogenic M2 phenotypes\textsuperscript{12}, while disease severity alone seems majorly affecting the numbers of M2 macrophages that feature in the pro-angiogenic process. Collectively, these results together showed that activated M2 macrophages caused by age and disease severity is the major immune cell type involved in the immune response in the pathogenesis of AMD.

Identification of age- and disease-severity-associated genes in relation to progression of AMD

We next investigated whether retinal aging is related to MGS level in AMD patients. Using the OLR model that controls sex and age, we correlated age-associated genes to MGS level, resulting in 50 genes (hereafter referred to Age-MGS genes, \( p < 0.05 \)) that their expression levels were significantly altered with the change of MGS level (Fig. 2, Extended Data Fig. 6, Supplementary Data 5). Among these Age-MGS genes, expression levels of 7 genes (including \textit{CYP26B1}, \textit{DHRS3}, \textit{MATN1}, \textit{RPL15}, \textit{AL136454.1}, \textit{CXCL11}, and \textit{CYP26A1}) were reduced with disease severity, while expression levels of the rest of the genes were escalated with disease severity. Interestingly, none of the identified Age-MGS genes were previously reported as AMD-associated genes in genome association studies. We then explored the biological processes of each Age-MGS gene involved by analyzing their corresponding Gene Ontology (GO) terms (Supplementary Data 6). We found that nearly half of Age-MGS genes are involved in the activation of immune cells or inflammatory pathways (Supplementary Data 6), indicating a strong immune response and inflammation occurred during the development of AMD over increased age. Other Age-MGS were clustered to groups with other biological processes, such as retinal endothelium-mediated angiogenesis and retinoic acid metabolic process.

Although none of the Age-MGS genes we identified has been previously reported in genome association studies in AMD, many plays critical roles in immune responses and
inflammation in the retina. For example, S100 calcium binding protein B (S100B), a well-known biomarker of active neural distress that is involved in many neurodegenerative diseases\(^\text{19}\), are pro-inflammation and can induce retinal degeneration in animal models\(^\text{20, 21, 22}\). L-selectin (SELL), a type-I transmembrane glycoprotein and leukocyte-expressed cell adhesion molecule\(^\text{23}\), mediates retinal inflammation by selective recruitment of leukocytes subsets, triggering adhesion, diapedesis and migration of inflammatory cells\(^\text{24}\).

Retinal endothelium plays important roles in maintenance of retinal statics particularly in the development of retinal vasculature. Among 50 Age-MGS genes, a cluster of genes are related to endothelium or angiogenesis (Supplementary Data 6), including apelin receptor (APLNR), fibroblast growth factor 1 (FGF1), matrilin 1 (MATN1), V-set and transmembrane domain containing 4 (VSTM4), fibulin 5 (FBLN5), myelin protein zero like 3 (MPZL3), and serine carboxypeptidase 1 (SCPEP1), and their expression (except for MATN1) were increased with the disease severity, indicating the involvement of endothelial cells throughout the different stage of AMD. APLNR expression is limited in the veins and proliferative endothelial cells during physiological retinal angiogenesis\(^\text{25}\), suggesting endothelial cells might be active at the early stage of AMD. Indeed, using APLNR inhibitors can selectively prevent pathological retinal angiogenesis in an oxygen-induced retinopathy mouse model\(^\text{25}\). Studies have shown FGF1 plays a role in retinal neovascularization through organization of various angiogenic pathways and coordination of cell-cell interactions\(^\text{26, 27}\). We found that MATN1 expression was decreased with the disease severity. This result is consistent with a study that showed increased angiogenesis during fracture healing in a MATN1 deficient mouse, and purified MATN1 proteins enabled inhibition of angiogenesis both \textit{in vitro} and \textit{in vivo}\(^\text{28}\). Of note, VSTM4, FBLN5, MPZL3 and SCPEP1, have been reported to promote sprouting angiogenesis by endothelial cells\(^\text{29, 30, 31, 32}\), while their role in retinal angiogenesis remains unstudied.

We also found that a group of Age-MGS genes are related to the regulation of retinoic acid metabolic process (Supplementary Data 6), including cytochrome P450 family 26 subfamily A member 1 (CYP26A1) or B member 1 (CYP26B1), dehydrogenase/reductase 3 (DHRS3) and serine carboxypeptidase 1 (SCPEP1), and their expression (except for SCPEP1) were reduced with disease severity. Retinoic acid is an active metabolite of vitamin A (retinol), an essential signaling molecule involved in the normal development of the ventral retina and optic nerve\(^\text{33}\) as well as maintenance of the blood-retinal barrier\(^\text{34}\). Dysregulation of retinoic acid can cause degenerative retinal diseases. CYP26A1, a retinoic acid catabolizing enzyme, was reported to be enriched in the fovea in the human retinas and functions for
photoreception in rod-free zone generation\textsuperscript{35}. Reduction of \textit{CYP26A1} may affect degradation of retinoic acid, ultimately causing dysfunctionality of photoreceptors. Similarly, \textit{CYP26B1} is also responsible for retinoic acid metabolism\textsuperscript{36}, however, its role in the retina has not been investigated yet. \textit{DHRS3}, a retinal reductase and a possible gene target of all-trans-retinoic acid, has been reported to be reduced in the liver during inflammation likely resulting in the perturbation of the whole-body vitamin A metabolism that occurred in conditions with inflammatory stress\textsuperscript{37}. Together, these results imply that reduction of genes associated with retinoic acid-metabolism may disrupt functionality in photoreceptors, a typical feature seen in AMD patients\textsuperscript{38}.

Some Age-MGS genes were not clustered into the three groups above mentioned (Supplementary Data 6) but their significance in relation to disease severity remained high (Fig. 2). Few studies investigated function of Monooxygenase DBH like 1 (\textit{MOXD1}), however, GO annotation of \textit{MOXD1} revealed its involvement in octopamine biosynthetic process, dopamine catabolic process, and norepinephrine biosynthetic process (Supplementary Data 6). Our results showed an increased expression level of \textit{MOXD1} with disease severity, implying a progressive elevation of those neurotransmitters in the retina of AMD patients with age and disease severity. Increasing evidence showed that neural inputs, such as chronic stress or inflammation, resulted in the release of neurotransmitters (e.g. norepinephrine) at specific vessels in the blood-brain barrier, that in turn augment the expression of chemokines in the endothelium to establish gateways through which immune cells can reach the central nervous system\textsuperscript{39,40,41,42}. Given that the retina displays similarities to the brain and spinal cord in terms of functional, anatomic and immunological properties\textsuperscript{43}, it is possible that neurotransmitters in the retina might be also increased to stimulate immune responses under chronic stress resulting from AMD. Type III intermediate filament proteins including glial fibrillary acidic protein (\textit{GFAP}), vimentin, desmin, peripherin and syncoilin (\textit{SYNC}, ranked No. 2 in the Age-MGS gene list according to \textit{p} value, Fig. 2), have been reported in association with retinal gliosis\textsuperscript{44}, a feature of AMD. A quantitative study in particular investigated the role of \textit{GFAP} expression at early and late stage of AMD and showed that upregulation of \textit{GFAP} in the inner retina was predominantly related to drusen, while increased expression of \textit{GFAP} in outer retina was associated with disruption of the RPE and blood-retinal barrier\textsuperscript{45}. This result suggested that other type III intermediate filament proteins, such as \textit{SYNC}, may be involved in the gliotic response in the retina subjected to chronic stress.
Correlation of Age-MGS genes and their enrichment in retinal cell types

We next sought to understand the correlation between the expression of Age-MGS genes. Correlation analysis revealed 3 clear clusters of Age-MGS genes with correlated expression using the EyeGEx database with sex and age controlled. (Fig. 3a). For ease of description, these clusters were numbered as cluster A, B and C. Cluster A consists of zinc finger protein 417 (ZNF417), AC092279.2, MPZL3 and TATA-box binding protein associated factor, RNA polymerase I subunit A (TAF1A), most of which are relevant to regulation of transcription or gene expression (Supplementary Data 6). Cluster B contains 38 Age-MGS genes that are involved in various biological pathways. Of which, complement C1s (C1s), major histocompatibility complex (MHC) class I-related (MR1) and solute carrier family 9 member A9 (SLC9A9) have stronger correlation (|R| ≥0.4) with other genes in this cluster, suggesting their potential role in coordinating other genes for multiple biological processes in AMD. C1s is a single chain glycoprotein and cleaves the components C4 and C2 to initiate the complement cascade when it becomes activated\(^46\). Transcriptome profiling of retina tissue from AMD donor eyes (31 normal and 26 AMD) suggested a direct role of dysregulation of C1s among other complement pathways in retinal degeneration\(^47\). Moreover, upregulation of C1s along with C4a and C2 were found in retinas of rats subjected to photo-oxidative damage\(^48\). However, the functional role of C1s in retinal degeneration remains unclear. MR1 molecule is distinct from MHC class I molecules that can present microbial riboflavin metabolites to a specialized subset of T cells known as mucosal-associated invariant T (MAIT) cells\(^49\). Interestingly, a study showed that MR1 was up-regulated at early stages of retinal degeneration in DBA/2J mice featured an immune response in the iris and death of retinal ganglion cells\(^50\), suggesting its potential role in the progression of immune-relevant retinal diseases. SLC9A9 is highly expressed in the brain and is implicated with immune regulation and endocytosis via mTOR signaling and cell survival\(^51\). However, little is known about its role in the retina. Cluster C consists of all down-regulated genes with disease severity, including MATN1, RPL15, CYP26A1, CYP26B1 and DHRS3, which are associated to regulation of retinoid metabolic process, extracellular matrix organization, and cytoplasmic translation (Supplementary Data 6). It is unclear how these molecules interact in the progression of AMD.

To assess the cell type-specific expression pattern of the age-associated genes and Age-MGS genes, we further analyzed the neural retina single-cell RNA-seq (scRNA-seq) data from Human Cell Atlas (HCA, Extended Data Fig. 7)\(^52\). We found that the expression pattern of the age-associated genes is distinctive among 9 retinal cell types, and a higher
percentage of cells expressing these genes are Müller glia, microglia, ganglion and astrocytes (Extended Data Fig. 8). Age-MGS genes are more enriched in Müller glia, microglia and astrocytes (Fig. 3b). Müller glia and astrocytes express a few similar Age-MGS genes but their percentage of expressing cells are clearly different (Fig. 3b). Consistent with our finding, a previous study reported that Müller glia and astrocytes among other major retinal cell types express the highest number of AMD-associated genes identified by a GWAS study\(^53\). Müller glia and astrocytes assist metabolism of retinal neurons through releasing trophic factors, recycling neurotransmitter glutamate, and controlling extracellular ion homeostasis\(^54, 55\), while astrocytes are resident macrophages and are implicated in neuroinflammatory changes during age-dependent retinal diseases, including AMD\(^56\). Together, these analyses highlight the cellular landscape of Age-MGS genes in the context of retinal cell types in which these genes are normally expressed. Of note, many Age-MGS genes were not expressed in any of the neural retina cell types, and their expression in retinal cell types in the condition of AMD is not clear. We also found that correlation between AMD-associated genes identified through different methodologies, such as LRT combined with OLR (present study), transcriptome-wide association study (TWAS using EyeGEx)\(^7\) and GWAS\(^4\), is clearly distinct (Extended Data Fig. 9a), indicating that these is limited interaction between these genes. However, genes identified by TWAS (9 genes) and GWAS (25 leading genes from 34 risk loci associated with AMD, sample size >15,000) are closer to each other as analyses of TWAS are based on GWAS result. Using HCA data, we validated leading AMD-associated genes previously identified GWAS\(^4\) and showed similar results that higher percentage of associated cells expressing AMD-associated genes are Müller glia and astrocytes (Extended Data Fig. 9b).

**Functional roles of Age-MGS-associated genes in AMD**

To determine the functional roles of these Age-MGS genes involved in the pathogenesis of AMD, we applied functional enrichment analysis (Biological Process) by Metascape. The results revealed significant enrichment most related to regulation of innate immune response (Fig. 4a, Supplementary Data 7). Among which, innate immune response, leukocyte activation, and multicellular organismal homeostasis are centered in the cluster of GO terms and have most interaction (n ≥5) with other GO terms (Fig. 4b), suggesting their important roles in regulation of Age-MGS genes. Therefore, we combined genes involved in these biological processes (Supplementary Data 7) and focused on the 8 Age-MGS genes with significant changes with both MGS levels (Fig 4c) and age (Fig 4d), including C1s, C7, MR1,
phospholipase C gamma 2 (PLCG2), tripartite motif containing 22 (TRIM22), transmembrane protein 98 (TMEM98), CYP26B1, and VSTM4.

C1s and MR1 may play a central role in the regulation of innate immune responses because these genes have higher correlation scores with more Age-MGS genes that we aforementioned (Fig. 3a) and are reported to be related to retinal degeneration. However, their individual role in regulating immune response in the retina remains unclear. Complement C7 as one of terminal proteins of complement pathway including C5, C6, C8 and C9 involves complement-mediated inflammation in the retina in AMD and is protective for AMD development in Caucasians. These findings provided further evidence for complement overactivation in AMD pathogenesis. An integrating analysis using in silico genetic and biological pathway data showed that PLCG2 is associated with AMD based solely on genetic burden. Our study in line with this finding showed a progressively increased expression level of PLCG2 in the retina of AMD patients with age and MGS level. Both TRIM22 and TMEM98 are involved in biological processes such as immune response or inflammation (Supplementary Data 6), while their function in the stressed retina is not clear.

CYP26B1 as one of few down-regulated aging genes with disease severity play a role in regulation of retinoid metabolic process, while VSTM4 contributes to the stability of retinal vasculature (Supplementary Data 6). Together, our results suggested that these identified Age-MGS genes collectively regulate innate immune response, promote leukocyte activation and endothelial cell proliferation and ultimately retinal angiogenesis.

The cellular landscape of the retina of patients with AMD and its relation to Age-MGS-associated genes

Previous studies have shown cellular changes occurred in the retina of patients with AMD, including Müller glia, astrocytes, microglia and retinal pigment epithelium. However, detailed quantification of cellular changes remains a challenge. Here we deconvoluted the bulk retina EyeGEx transcriptomes with CIBERSORTx to quantify cellular changes, using the scRNA-seq data from HCA as a reference (Extended Data Fig. 7, Supplementary Data 8). To increase the confidence of down-stream analyses, the cell types were retained if their estimated proportions were >0 in all samples (Extended Data Fig. 10).

To assess whether age as a risk factor affects cell type proportions, we used an OLR model to control for sex and MGS level and identified age-associated alteration in the proportions of retinal cell types. We observed that the proportions of Müller glia were significantly increased with age, while the proportions of horizontal and rod cells were
significantly decreased with age (Fig. 5a, Extended Data Fig. 11, Supplementary Data 9).

Next, we focused on the risk factor of MGS level to assess its effect on cell proportions by controlling for sex and age using an OLR model. Although the proportions of Müller glia and rod were increased and decreased, respectively, with MGS levels, the changes were not significant (Fig. 5b, Extended Data Fig. 12, Supplementary Data 10), suggesting that disease severity has limited effects to alter cell proportions in the retina of patients with AMD. Of note, ganglion cells were removed in Fig. 5b due to its zero estimated expression profile across all samples after adjustment for sex and age. Previously study using immunolabelling of glial fibrillary acidic protein (GFAP), a marker of retinal glial cells, observed morphological changes and upmodulation of astrocytes and Müller glia in the inner and outer retina, respectively in a small number of human donor retinas (normal=6, AMD=11)\(^{45}\). Rods loss in ageing and age-related maculopathy has been widely reported\(^{62, 63}\). Moreover, Rods die in older eyes without evidence of overt retinal pigment epithelial degeneration\(^{62}\), indicating that age is a main factor to cause rods loss. Horizontal cells form triad synapses with photoreceptors and bipolar cells to module signal transmission in between these cells\(^{64}\).

scRNA-Seq profiling of the aging human retina revealed a horizontal cell subtype (ISL1\(^{+}\) CALB1\(^{+}\)) were greatly reduced particularly in fovea, suggesting that age may affect different regions in the retina\(^{65}\). Together, these results suggested that age is a central risk factor in AMD that enables alterations in the proportions of cell types in the retina, i.e. Müller glia, horizontal cells and rod photoreceptors.

As most of Age-MGS genes are enriched in retinal glial cells including Müller glia and the increased proportions of Müller glia with age, we specifically investigated the relationship between Müller glia and 2 core Age-MGS genes we identified (Fig. 5c). We showed that expression of \(C1s\) and \(MR1\) have a strong positive (Pearson correlation coefficient, \(0.5 < r < 1\)) relationship with the estimated proportions of Müller glia. We also observed strong negative (-0.5 \(< r < -1\)) correlation between expression of \(C1s\), \(MR1\) and the estimated proportions of rod photoreceptors (Extended Data Fig. 13a), while no significant or weak positive correlation (\(0 < r < 0.3\)) between these two genes and the estimated proportions of horizontal cells (Extended Data Fig. 13b). These results suggested that \(C1s\) and \(MR1\) are potential core Age-MGS genes to regulate Müller glia and rod photoreceptors but not horizontal cells that are involved in retina aging. Notably, among the three cell types there is a relatively higher percentage of Müller glia expressing \(C1s\) and \(MR1\) (Fig. 3b). Photoreceptor degeneration and Müller glia activation are indicative of structure and molecular abnormalities in the development of AMD\(^{66}\), suggesting their possible complementary
relationship in this form of retinal degeneration. Indeed, our results revealed that there is a strong negative correlation (r = -0.89) between rod photoreceptors and Müller glia based on its estimated proportions after controlling for sex and age (Fig. 5d).

**Discussion**

In the present study, we systematically analyzed the retinal aging transcriptome and cellular landscape in association with the progression of age-related macular degeneration. We revealed that the aging retina is characterized by a range of biological processes, in particular innate immune response and inflammation that could lead to the progressively severe AMD. We specifically studied the immune cell profiles of the EyeGEx database and identified the upregulation of M2 macrophages and downregulation of M0 cells with both age and disease severity in AMD patients. We subsequently pinpointed a group of retinal aging genes that change with disease severity (Age-MGS genes) by controlling for sex and age. Many of the Age-MGS genes have function mainly in relation to immune response, endothelium-mediated angiogenesis and retinoic acid metabolic process. Through gene correlation analysis, we demonstrated that \textit{C1s}, \textit{MR1} and \textit{SLC9A9} have stronger correlation with other Age-MGS genes, and all are involved in regulation of immune response. By integrating these data with the single-cell transcriptome of the human neural retina, we further identified specific cell types that normally express these Age-MGS genes. GO enrichment analysis of the Age-MGS genes further suggested the aging and AMD retina is characterized by innate immune response, leukocyte activation and multicellular organismal homeostasis. We further identified 8 Age-MGS genes in these biological processes that potentially play critical roles in the progress of AMD development. Through bulk deconvolution analyses, we showed that the proportions of Müller glia are increased with age, while the proportions of rod photoreceptors and horizontal cells are decreased with age. Moreover, the proportions of all retinal cell types do not have significant changes with disease severity after factoring out sex and age, suggesting that age is potentially the central risk factor that uniquely contributes to the severity of AMD. Correlation analysis between 8 Age-MGS genes and the proportions of cells with significant alteration revealed that \textit{C1s}, \textit{MR1} and \textit{TMEM98} have a strong relationship with Müller glia and rod photoreceptors, both of which are negatively correlated. Future research to investigate the function of these Age-MGS genes in the retina would be important to understand the mechanism underly AMD pathology.
Since the transcriptome data from human donors are complexed with a variety of latent factors, such as sex, age and clinical diagnosis, it is important to use batch correction to adjust those factors for the downstream analysis. Throughout our analysis, all the raw data were subjected to the batch correction before relevant analysis. We noted a study using various deconvolution algorithms to analyze involvement of retinal cell types in AMD patients with different databases, including EyeGEx\textsuperscript{67}. However, all the raw data from this study was not subjected to batch correction, thus the conclusion resulted from was less stringent compared to the present study. A clear example of difference between adjusted and unadjusted data analysis can be referred to the cell type proportion estimated by CIBERSORTx (Extended Data Fig. 3, 5, 11 and 12). There were several limitations of this study. For examples, the sample size in each of age group is relatively small particularly the group of 100-109, which likely affects the number of aging genes defined by the LRT model. Moreover, the HCA dataset does not include RPE cells for which we might miss information on how the Age-MGS genes involving in the degeneration of RPE cells in AMD\textsuperscript{68}. In addition, we used the human hematopoietic immune cell matrix, LM22, to deconvolute immune cell profiles in the retina, potentially missing information on resident immune cell types, such as microglia, in neural tissues including brain and retina. Future studies to determine the change of the expression of Age-MGS gene in RPE cells and to generate transcriptome dataset for resident immune cells in neural tissues would be helpful. Our study defined the aging genes in the retina using the LRT model by controlling for sex and disease severity of AMD and identified novel ageing genes associated with AMD progression. We also highlighted the potential for using deconvolution analysis to understand cellular and transcriptomic changes in AMD. The functions of most of the Age-MGS genes remain unclear in the retina and more research is needed to understand their role in progressive development of AMD. Our analyses provide potential directions for subsequent research on the aging retina for which AMD progressively gets worse over age.

Methods

Data accession

RNA-Seq raw counts of the Eye Genotype Expression (EyeGEx) database were obtained from Gene Expression Omnibus (GSE115828, accessed on 11\textsuperscript{th} January 2022)\textsuperscript{7}. Raw counts were converted to transcripts per million (TPM) in R (v4.1.0) for downstream analyses. Quality-controlled samples (105 MGS1, 175 MGS2, 112 MGS3, and 61 MGS4 samples)
were selected from the raw dataset using the subset function from the R package Seurat\textsuperscript{69} (v4.0.5). Detailed clinical information including patient sex, age and MGS level were accessed from GSE115828. Single-cell transcriptomes of human neural retina were directly downloaded from Human Cell Atlas (https://www.humancellatlas.org/, E-MTAB-7316, accessed on 15\textsuperscript{th} January 2022)\textsuperscript{52}.

Identification of age-associated genes in human neural retina from donors with AMD

To identify age-associated genes, the RNA-Seq raw count from quality-controlled samples was analyzed by DESeq\textsuperscript{2}\textsuperscript{8}(v1.32.0), using the likelihood-ratio test (LRT). According to the design of batch correction in EyeGEx database, gene expression was determined by three variables, including sex, age and MGS level\textsuperscript{7}. Thus, sex, age and MGS level were included in the LRT model for controlling these factors. Age was categorized into decades, including 50-59, 60-69, 70-79, 80-89, 90-99, and 100-109, for downstream analyses. Significant age-associated genes were determined by adjusted $p < 0.05$. The resulting genes were then scaled to z-score and clustered using the degPatterns function from the R package DEGreport (v1.28.0). Gene clusters with consistent and progressive changes (increasing or decreasing with age) were selected for downstream analyses.

Enrichment analyses for selected genes were performed using Metascape (https://metascape.org/gp/index.html#/main/step1), a web-based tool that involves a variety of terms including GO Biological Processes\textsuperscript{9}. Importantly, Metascape provides more frequently updated bioinformatics analyses than DAVID\textsuperscript{70}.

Identification of age-MGS-associated genes in human neural retina from donors with AMD

To identify age-MGS associated genes, the TPM values of age-associated genes from both clusters (increasing or decreasing with age) were analyzed by R package MASS (v7.3-54) and ordinal (v2019.12-10) using ordinal logistic regression (OLR) model\textsuperscript{71}. Using Cumulative Link Model function in ordinal\textsuperscript{71}, we controlled sex and age to screen input genes that are associated with MGS level. Age-MGS-associated genes were determined at a significance threshold of $p < 0.05$. The resulting genes were represented in a forest plot and ranked according to its beta coefficient.

For gene-gene correlation of Age-MGS genes, the raw counts were processed by Variance Stabilizing Transformation function in DESeq\textsuperscript{2}\textsuperscript{8}, followed by statistical adjustment.
for sex and age using the Remove Batch Effect function in limma \(^{72}\) (v3.48.3). The correlation heatmap was plotted using R package corrplot \(^{73}\) (v0.92).

**RNA-Seq gene expression visualization and statistical analyses**

For visualization of adjusted gene expression levels, the raw counts were processed by Variance Stabilizing Transformation function in DESeq2 \(^{8}\), followed by statistical adjustment for sex and MGS level or sex and age using the Remove Batch Effect function in limma \(^{72}\) (v3.48.3). Comparison of gene expression among different age groups or MGS groups are presented as Tukey boxplots, with interquartile range boxes and 1.5x interquartile range whiskers. Two-tailed Mann-Whitney test was used to assess pairwise comparisons in the plots, while Kruskal-Wallis test was used to perform statistical comparison across all the age groups or MGS groups. Such statistics were performed to confirm the results of age-associated genes generated by the two-sided DESeq2 LRT \(^{8}\).

**Normal human neural retina scRNA-Seq data analysis**

scRNA-seq raw counts from HCA were analyzed using Seurat \(^{69}\) (v4.0.5). Cells with >200 detected genes and expressing < 10% mitochondrial genes were retained, resulting in 19,950 cells. Genes were retained if they were expressed in ≥ cells. The LogNormalize method was used to normalise the dataset, followed by removing inherent variation caused by mitochondrial gene expression and the number of unique molecular identifiers per cell \(^{52}\). Using the graph-based shared nearest neighbour method, clustering at a resolution of 0.9 was performed on PCA-reduced expression data because it yields high maximum cluster assignment probability for most of the cells. Clusters were visualized using uniform manifold approximation and projection (UMAP). Retinal cell markers were referred to studies analyzing retina scRNA-Seq data \(^{52,53}\).

Of the 50 Age-MGS-associated genes identified from the EyeGEx transcriptome, 43 genes were matched in the HCA dataset. To calculate the proportions of cells expressing a given gene, the expression matrices were converted to binary matrices using the threshold of expression > 0 \(^{74}\). Expression frequencies of a given gene in different cell types were determined as per the cell type annotation resulting from above analyses. We further scaled the expression frequencies in cell types in z-score using the NMF package (v0.23.0).

**Deconvolution analyses of the immune cells and retinal cells from donors with AMD**
To infer the estimated proportions of infiltrating immune cells or retinal cells of each bulk retinal transcriptome, we used the CIBERSORTx algorithm, a deconvolution algorithm that uses a set reference gene expression value\(^ 1 \). For the immune cellular profile, we used the signature matrix, LM22, defining 22 infiltrating cells\(^ 35 \). Normalized data (TPM value) of the EyeGEx transcriptome were uploaded to the CIBERSORTx web portal (https://cibersortx.stanford.edu/), with the algorithm run using LM22 signature matrix at 100 permutations with B-mode batch correction.

For retinal cellular profile, we derived a signature matrix using HCA scRNA-Seq dataset. 10,000 retinal cells including rod, Müller glia, bipolar, cone, amacrine, microglia, ganglion, astrocytes, and horizontal cells, were acquired using the Seurat subset function and labelled with corresponding cluster identities. The labelled cells were used to establish a single-cell signature matrix with the following settings: Min.Expression (0.05), replicates (10), and sampling (0.5), kappa (999), q-value (0.01) and No. Barcode Genes (100-500). The resulting signature matrix was used to impute retinal cellular fraction from the EyeGEx transcriptome with S-mode batch correction. The number of permutations to test for significance were set at 100. Statistical significance of age- or MGS-level-association was assessed by an ordinal logistic regression model using the nonparametric Kruskal-Wallis test that allows for multi-factorial designs. Sex and MGS-level or sex and age were included in the regression model to reveal the cellular fractions that are significantly altered with age or MGS-level, respectively. The estimated coefficients were presented as a forest plot with 95% confidence intervals. The estimated proportions of cell types were adjusted for sex and MGS-level or sex and age for assessment of effects by factor of age or MGS-level by extracting the residuals from fitting a generalized linear model with variables.
Statistical analyses

Detailed information of statistical analyses was described in different places, including results, figures, and figure legends, in which the test methods, significance, \( p \) values, error bars and coefficients were included.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All processed data files generated for this study are included in this article and its supplementary information files. Raw data files from two sources are available for download through GEO under the accession number GSE115828 and Human Cell Atlas under accession number E-MTAB-7316.

Code availability

Analysis scripts will be made available upon request.

Acknowledgements

We thank Ryan D. Chow for sharing codes for data analyses at Github.

Funding

This work was supported by grants from the National Health and Medical Research Council of Australia (GNT1185600). The Centre for Eye Research Australia receives Operational Infrastructure Support from the Victorian Government.

Author contribution


Conflict of interests

The authors declare no competing interests.
References


**Figure legends**

**Fig. 1. Identification of age-associated genes in the human retina from donors with AMD.**

*Fig. 1.* Demographics of the human retina RNA-Seq profiles in the EyeGEx dataset (n=453, previously quality-controlled dataset), detailed by age group and AMD severity (MGS level).

*Fig. 1.* Tukey boxplots (interquartile range (IQR) boxes with 1.5× IQR whiskers) of age-associated genes in the human retina. Expression of Age-up genes increases with age (left, n=445), while expression of Age-down genes decreases with age (right, n=57). Age-associated genes were defined with DESeq2 two-sided likelihood-ratio test (adj. p < 0.05) by controlling for sex and, MGS level. Adjusted gene expressions were shown as z-score.

*Fig. 1.* A heatmap of expression level of all age-associated genes in all samples across different age group, adjusted for sex and MGS levels.

*Fig. 1.* Functional enrichment analyses (biological processes) of age-up genes by Metascape.

**Fig. 2. Identification of age-associated genes progressively altered with AMD severity (MGS level).**

Forest plot of age-associated gene having significant correlation with MGS level (p < 0.05, n=50). Positive coefficients indicate age-associated genes increase in expression with MGS level, while negative coefficients indicate age-associated genes decrease in expression with MGS level (i.e. Age-MGS genes). Statistical analyses of MGS association for age-associated genes were assessed by a nonparametric ordinal logistic regression model that controls for sex and age. Point sizes are scaled by statistical significance. Error bars represent 95% confidence intervals.

**Fig. 3. Correlation of Age-MGS genes and their enrichment in retinal cell types.**

Heatmap of the correlation matrix across 50 Age-MGS genes identified in Fig.2. Pearson’s correlation was calculated among 50 Age-MGS genes to show the co-expression patterns of genes in the heatmap. Colour key denotes the Pearson’s correlation higher than absolute value 0.3 between genes. There are 3 clear clusters (labelled as A, B and C) of Age-MGS genes with correlated expression adjusted by sex and age. Genes highlighted in red indicates their stronger correlation with many other genes.

*Fig. 3.* Heatmap showing the percentage of retina cells expressing each of Age-MGS genes, scaled by gene across the different cell types. Data are from the Human Cell Atlas (HCA).

**Fig. 4. Functional roles of Age-MGS genes in AMD.**

*Fig. 4.* Functional enrichment analyses (gene ontology annotations of biological processes) of 50 Age-MGS genes by Metascape.
Interaction network of gene ontology annotations described in Fig. 4a. Annotations highlighted in red indicates that the specific biological process has the most interactions with other biological processes. c, Tukey boxplots (interquartile range (IQR) boxes with 1.5× IQR whiskers) showing the expression of C1s, C7, MRL, PLCG2, TRIM22, TMEM98, CYP26B1, and VSTM4. Gene expression value are shown as log-transformed, controlled for sex and age (Fig. 4c) or sex and MGS level (Fig. 4d). Statistical significance of age-associated or MGS-associated difference was assessed by two-sided Kruskal-Wallis test on the adjusted expression values.

Fig. 5. The cellular landscape of the retina of patients with AMD and its relation to Age-MGS genes. a, b, Forest plot of estimated proportions of retinal cells having significant correlation with age or MGS level, respectively (p < 0.05). Positive coefficients indicate the proportion of cells increases with age or MGS level, while negative coefficients indicate the proportion of cells decrease with age or MGS level. Statistical analyses of age association or MGS association for the proportions of cells were assessed by a nonparametric ordinal logistic regression model, adjusted for sex and MGS level (Fig. 5a) or sex and age (Fig. 5b), respectively. Point sizes are scaled by statistical significance. Error bars represent 95% confidence intervals. c, Analyses of the correlation between expression level of C1s or MRL and the estimated proportion of Müller glia, adjusted for sex and age. Pearson’s correlation coefficient was used to test the strength of linear relationships between gene expression and the estimated proportion of Müller glia. Black dots denote individual samples (n=453). Blue lines denote regression lines. d, Heatmap of the correlation matrix among retinal cell types, based on the proportions of retina cells estimated by CIBERSORTX and adjusted for sex and age. Pearson’s correlation was calculated among retinal cells to show the co-expression patterns of cells in the heatmap. Values in the heatmap indicates the Pearson correlation coefficient between two cell types.
a) Retinal RNA-seq profiles from AMD patients and healthy controls (EyeGEx).

b) Age-up genes (n=445) and Age-down genes (n=57).

c) Heatmap showing gene expression across different age groups.

d) Heatmap showing enriched biological processes for age-up and age-down genes.