# Paired DNA and RNA sequencing uncovers common and rare genomic variants regulating gene expression in the human retina

Jacob Sampson[1], Ayellet V Segrè[2,3,4], Kinga M Bujakowska[2,3], Simon J Clark[1,5], Paul N Bishop[1], Steve Haynes[1], Diana Baralle[6], Jospin Al-Deek[1], Stacey Holden[1], Beverley Anderson[1], Andy Hayes[1], Rahmat A Kemal[1,7], Huw B Thomas[1], Raymond T O'Keefe[1], Siddharth Banka[1,8], Graeme C Black[1,8], Panagiotis I Sergouniotis[1,8,9,10], Jamie M Ellingford[1]

[1] Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

[2] Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear, Boston, MA, USA.

[3] Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

[4] Broad Institute of Harvard and MIT, Cambridge, MA, USA

[5] University Eye Clinic, Eberhard Karls University of Tübingen, Tübingen, Germany

[6] School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

[7] Department of Medical Biology, Faculty of Medicine, Universitas Riau, Pekanbaru, Indonesia

[8] Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK

[9] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL- EBI), Wellcome Genome Campus, Cambridge, UK

[10] Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, UK

Correspondence:

jamie.ellingford@manchester.ac.uk

# Abstract

Genetic disorders impacting vision affect millions of individuals worldwide, including age-related macular degeneration (common) and inherited retinal disorders (rare). There is incomplete understanding of the impact of genetic variation on gene expression in the human retina, and its role in genetic disorders. Through the generation of whole genome sequencing and bulk RNA-sequencing of neurosensory retina (NSR) and retinal pigment epithelium (RPE) from 201 post-mortem eyes, we uncovered common and rare genomic variants shaping retinal expression profiles. This includes 1,483,595 significant cis-expression quantitative trait loci (eQTLs) impacting 9,959 and 3,699 genes in NSR and RPE, respectively, with associated genomic variants enriched to cis-candidate regulatory elements and notable shared eGenes between NSR and RPE. We also detected 1051 expression outliers and prioritised 299 rare non-coding single-nucleotide, structural variants or copy number variants as plausible drivers for 28% of outlier events. This study increases understanding of gene expression regulation in the human retina.

# Introduction

Genomic variation has been well established to play a role in the onset and susceptibility of visual impairment by disrupting normal functioning of the retina, a highly specialised light-sensitive tissue at the back of eye (Wright *et al.*, 2010). The retina depends on the interaction between neuronal and non-neuronal cell types, including those in the neurosensory retina (NSR), e.g. photoreceptors and ganglion cells (Hoon *et al.*, 2014), and the retinal pigment epithelium (RPE), a monolayer which lines the photoreceptor outer segments (Strauss, 2005). Inherited retinal disorders (IRDs) are a diverse set of largely monogenic conditions driven primarily by highly impactful genetic variants that are rare in the population and disrupt function of the NSR and/or RPE (Hanany, Rivolta and Sharon, 2020). Monogenic IRDs may present in isolation, for example, Stargardt disease, retinitis pigmentosa and cone-rod dystrophy, or as part of a multi-system disorder, for example Usher syndrome, Joubert syndrome and Senior-Loken syndrome. Age-related macular degeneration (AMD) is a common disorder that impacts the retina and is a leading cause of visual impairment in adults (Fleckenstein *et al.*, 2021), predicted to impact 288 million individuals by 2040 (Wong *et al.*, 2014). Whilst non-genetic risk factors exist for AMD, including age, diet and lifestyle, its heritability is estimated to be as high as 71% (Seddon *et al.*, 2005). Genome wide association studies (GWAS) initially identified more than 50 genomic loci impacting 34 genes that convey high risk to AMD in a European ancestry cohort (Fritsche *et al.*, 2016). Recent expansion of AMD GWAS to Hispanic and African ancestries has uncovered 30 additional genomic loci and distinct AMD genomic architecture in these populations (Gorman *et al.*, 2024).

The Genotype-Tissue Expression (GTEx) project has transformed our ability to pinpoint genetic variants that impact gene expression (THE GTEX CONSORTIUM, 2020), including tissue-specific and tissue-shared expression quantitative loci (eQTLs) and rare genetic variants associated with expression outlier (eOutlier) events. Findings from these investigations, along with other studies, have been leveraged across various biological and medical fields to gain a deeper understanding of disease mechanisms (Barbeira *et al.*, 2021; Hamel *et al.*, 2024), to provide more informed diagnosis and prognosis (Michaud *et al.*, 2022), and to pursue pathways for novel treatments (Davenport *et al.*, 2018). Notably, ophthalmic tissue was not included in the GTEx resource, but recently, the EyeGEx study identified over 2 million eQTLs in the retina, from a cohort comprised of healthy eyes and eyes displaying signs of AMD (Ratnapriya *et al.*, 2019) and (Strunz *et al.*, 2020) identified 580,171 eQTLs in the neural retina from a cohort of healthy eyes. These resources enable the investigation of the role of common single-nucleotide variants (SNVs) influencing retinal gene expression. However, to our knowledge, there are no suitable datasets to also interrogate the impact of rare variants, structural variants (SVs) and copy number variants (CNVs) on retinal gene expression. Expanding our understanding of gene expression regulation in the retina will provide insights into the molecular mechanisms underlying both common and rare eye diseases, and help identify new potential strategies for treatment and prevention.

Here we describe the creation of a unique resource of paired genomes and transcriptomes for the human retina from 201 donors, and develop new understanding of both common and rare variants which drive expression in this highly specialised tissue.

# Results

## The METR genome-transcriptome resource integrates genomic and retinal transcriptomic data from 201 post-mortem eye samples

The Manchester Eye Tissue Repository (METR) genome-transcriptome cohort is comprised of 201 unrelated individuals that donated eye tissue post-mortem. The median age of the cohort was 71 years (IQR 64-77) at time of post-mortem, with a slight male predominance (63.7%). The median ischemic time was 40 hours (IQR=32-44) ***(Supplementary figure 1***). While 47 individuals (23% of the cohort) were found to carry genetic variants that confer high-risk for age-related macular degeneration (AMD), none of the 201 individuals included in the cohort had phenotypic presentation, assessed post-mortem, consistent with late-stage AMD or monogenic ophthalmic disorders.

Short-read whole genome sequencing was performed on an Illumina NovaSeq6000, with alignment and variant detection performed using DRAGEN software (v4.0.3). The median genome-wide average coverage per sample was 35.9x (IQR=30.3–40.5) (***Supplementary figure 2***), with an average of 88.0% and 92.8% of the genome covered by at least 15 or at least 10 sequencing reads, respectively. Joint SNV calling with DRAGEN PopGen 4.2.4 obtained aggregate calls at 15,617,784 high-confidence variant sites after quality control (***Supplementary table 1***). On average, 173 CNVs and 8,814 SVs were identified per sample (***Supplementary table 2***). Genomic variation profiles among the 201 donors confirmed the cohort was exclusively of European genetically inferred ancestry (***Supplementary figure 3***).

Transcriptomic data was generated by short read bulk-RNA sequencing of polyadenylated enriched RNA, using an Illumina NovaSeq6000. The median RNA integrity number (RIN) for samples selected for transcriptomic analysis was 7.9 (IQR=7.5-8.1) for NSR samples (n=183) and 6.9 (IQR=6.5-7.5) for RPE samples (n=176) (***Supplementary figures 1 and 4A***, ***Supplementary table 3***). We obtained an average of 139 million uniquely mapped reads for NSR (IQR=138-161 million) and 62.3 million for RPE (IQR=59-94 million), representing 89.4% (IQR=88.2-90.7%) and 82.9% (IQR=78-86%) of all generated reads, respectively (***Supplementary figure 4***). The median 3'/5' bias, defined as the ratio of sequencing depth between the 150 bp region at the 3' end and the 5' end of the gene for genes with a length greater than 600 bp and at least 5 unambiguous reads, was 0.5 for NSR samples (IQR=0.48-0.51) and 0.51 for RPE samples (IQR=0.49-0.55) (***Supplementary figure 1B***). Some level of expression (mean TPM > 0.1) was indicated for 28,512 genes across both tissues, with 18,891 and 13,214 genes expressed at moderate (TPM >1) and high (TPM > 5) levels, respectively (***Figure 1A***). 59% of expressed genes (mean TPM > 0.1) (n = 16,785) and 90% of highly expressed genes (mean TPM > 5) (n = 11,984) were protein coding, representing 84% and 60% of all GENCODE protein coding genes, respectively. Significantly higher expression variability was observed in the RPE compared to the NSR for genes expressed at low, moderate and high levels in samples with both NSR and RPE data, measured by the coefficient of variation (p-value<$2.2\times10^{-16}$; ***Figure 1B***).

To ensure the validity of the transcriptomic datasets generated in this study, we assessed the biological relevance of expressed genes in NSR and RPE. Gene expression profiles were enriched for gene ontology (GO) terms indicative of the tissues of origin (***Figure 1C***). Overall, 14,957 differentially expressed genes (mean expression > 1 TPM and adj. p-value < 0.05) were identified between the RPE and the NSR. Unsurprisingly, cell type deconvolution analyses, with reference to single-cell retinal datasets (Monavarfeshani *et al.*, 2023), demonstrated a significantly higher representation of RPE cells in data generated from RPE samples compared to data generated from NSR samples (***Supplementary figure 5A***). Moreover, genes with increased expression in the RPE (n = 7,353) were enriched for 987 GO terms, which were grouped into 55 non-redundant clusters, including epithelial cell proliferation (GO:0050678), regulation of cell adhesion (GO:0030155) and positive regulation of immune response (GO:0050778) (***Supplementary table 4***). Deconvolution of the NSR datasets supported the presence of at least 7 neuronal cell types at high levels (estimated proportion > 1%), with an average relative composition, per sample, of 29% rod photoreceptors (95%CI:26.4-30.7%), 28% retinal astrocytes (95%CI=26.1-29.6%), 16% amacrine cells (95%CI=15.7-17.1%), 10% horizontal cells (95%CI=9.5-10.2%), 7% retinal ganglion cells (95%CI=5.8-9%), 4% bipolar cells (95%CI=3.8-4.4%), 2% Müller glia (95%CI=1.7-3.1%) and ~4% other cell types *(**Supplementary figure 5C**).* Genes with increased expression in the NSR, compared to the RPE (n = 7,604), were enriched for 238 GO terms, grouped into 27 non-redundant clusters including synapse organisation (GO:0050808), neurotransmitter transport (GO:0006836) and cell morphogenesis involved in neuron differentiation (GO:0048667) *(**Supplementary table 4**)*.

## METR eQTLs provide novel insights into non-coding variants that impact known eye disease-related genes

We performed cis-eQTL mapping to identify common genetic variants that are associated with gene expression in the NSR and the RPE. We found 1,424,946 significant (FDR < 0.05) cis-eQTL associations between 806,789 variants (eVariants) and 9,959 genes (eGenes) in NSR ***(Supplementary figure 6).*** Additionally, 465,045 eQTLs were identified between 303,773 eVariants and 3,699 eGenes in the RPE *(**Supplementary figure 6**)*. The lower range of alternate internal allele frequencies for eVariants identified as part of eQTLs was 2.5% and included novel variants in comparison to gnomAD v4.1 (***Supplementary figure 7***). 406,396 eQTLs were common to both the retina and the RPE, while 1,018,550 associations were NSR-specific and 58,649 were RPE-specific *(**Figure 2A**).* Henceforth, we will refer to eQTLs identified in the NSR and/or the RPE as METR-eQTLs (n=1,483,595), which included 10,471 unique eGenes (6,772 NSR-specific, 512 RPE-specific and 3,187 eGenes in both NSR and RPE).

We compared the top eQTLs identified for NSR (METR-NSR eQTLs) for each eGene identified in this study with two published retina-specific eQTL datasets, the EyeGEx project (Ratnapriya *et al.*, 2019) and (Strunz *et al.*, 2020), to identify: (1) eQTLs identically replicated in the NSR tissues; (2) METR-NSR eQTLs that impacted eGenes previously described but had alternative eVariants in high linkage disequilibrium (LD) with findings from EyeGEx or

Strunz et al. (2020), and (3) previously unreported eQTLs for NSR, including newly identified eGenes *(**Figure 2B; Supplementary figure 8**)*. Of note, our cohort excludes individuals with late-stage AMD, whereas EyeGEx includes late-stage AMD eyes. We report 6,181 NSR eGenes which were previously described by at least one other study (62% of all NSR eGenes), of which 547 eGenes (9%) share identical top eVariants with at least one other study and 1,882 (30%) have top eVariants in high LD with previously identified top eVariants ($r^2$> 0.8) (*Figure 2B*). We identified 343,527 novel eQTLs (24%) in eGenes that were previously described by at least one previous study and 386,741 novel eQTLs (27%) in 3,778 newly identified eGenes. Importantly, we replicate 13 eQTLs that have previously been reported to impact genes that are implicated in increased risk of AMD (***Table 1***).

***Table 1 Replication of eQTLs that impact AMD risk genes and were identified as lead candidates by Ratnapriya et al. (2019) and Orozco et al. (2020) in the METR-eQTL dataset.*** *26 eQTLs across 14 AMD risk loci have previously been identified. Of these, 15 were either replicated (n = 13) or in high LD (n = 2) ($r^2$>0.8) with METR-eQTLs. Novel METR-eQTLs are identified for 5 genes previously implicated in AMD risk. 10 of the 21 (48%) unique eVariants overlap with characterised candidate cis-regulatory elements, although only 1 of these has previously been characterised as active in retinal cells. Definitions utilised in table: 'Replicated', lead candidate variant from other studies is also identified in this study; 'METR eQTL in high LD', lead candidate variant from other studies is in high LD with an eQTL identified in this study; 'Novel METR eQTL', the lead candidate variant from other studies is not replicated in this study, additional and previously unreported eVariants are identified in this study but they are not in high LD with the previously reported lead candidate variant; 'No METR eQTL', the gene is not identified as an eGene in this study within the statistical parameters applied (FDR<0.05); 'Not replicated', we identify eQTLs for the gene in this study but the lead candidate variant from other studies is not replicated nor is it in high LD with an eQTL identified in this study.*

| AMD risk locus | Gene ID | Reported candidate eVariant from previous study | Source | Category | Overlapping regulatory elements (tissue; type) |
|---|---|---|---|---|---|
| ACAD10 | SH2B3 | chr12_111694806_G_A | EyeGex | Novel METR eQTL | |
| ARMS2/HTRA1 | HTRA1 | chr10_122534138_C_T | Orozco *et al.* | Novel METR eQTL | |
| B3GALTL | B3GLCT | chr13_31247103_C_T | EyeGex | Not replicated | Non-retina; enhancer |
| B3GALTL | B3GLCT | chr13_31133338_T_A | Orozco *et al.* | Novel METR eQTL | |
| CFI | PLA2G12A | chr4_109737911_T_C | EyeGex | Novel METR eQTL | |
| CFI | CFI | chr4_109594995_C_G | Orozco *et al.* | No METR eQTL | |
| COL4A3 | COL4A3 | chr2_227254233_C_G | Orozco *et al.* | No METR eQTL | |
| CTRB2/CTRB1 | BCAR1 | chr16_75208831_T_C | Orozco *et al.* | METR eQTL in high LD | |
| MMP9 | SLC12A5-AS1 | chr20_46006318_C_T | Orozco *et al.* | Replicated | Non-retina; enhancer; |
| NPLOC4-TSPAN10 | TSPAN10 | chr17_81654604_A_G | Orozco *et al.* | Replicated | |
| PILRB/PILRA | PILRB | chr7_100393925_C_T | EyeGex | Replicated | Retina and non-retina, enhancer, scATACseq (cones, bipolar cells |
| PILRB/PILRA | PILRA | chr7_100393925_C_T | EyeGex | Replicated | Retina and non-retina, enhancer, scATACseq (cones, bipolar cells) |
| PILRB/PILRA | ZCWPW1 | chr7_100393925_C_T | EyeGex | Replicated | Retina and non-retina, enhancer, scATACseq (cones, bipolar cells) |
| PILRB/PILRA | TSC22D4 | chr7_100393925_C_T | EyeGex | Replicated | Retina and non-retina, enhancer, scATACseq (cones, bipolar cells) |
| PILRB/PILRA | PILRB | chr7_100375779_G_A | Orozco *et al.* | Replicated | |
| PILRB/PILRA | PILRA | chr7_100345960_A_T | Orozco *et al.* | Replicated | Non-retina; enhancer |
| RDH5/CD63 | BLOC1S1 | chr12_55721994_C_A | EyeGex | Replicated | Non-retina; enhancer |
| RDH5-CD63 | BLOC1S1 | chr12_55819513_G_T | Orozco *et al.* | Replicated | |
| RDH5-CD63 | RDH5 | chr12_55721801_C_T | Orozco *et al.* | No METR eQTL | Non-retina; enhancer |
| SLC16A8 | BAIAP2L2 | chr22_38071777_G_C | Orozco *et al.* | No METR eQTL | Non-retina; enhancer |
| TMEM97/VTN | POLDIP2 | chr17_28322698_A_C | EyeGex | Replicated | Non-retina; enhancer |
| TMEM97/VTN | SLC13A2 | chr17_28322698_A_C | EyeGex | Novel METR eQTL | Non-retina; enhancer |
| TMEM97/VTN | TMEM199 | chr17_28322698_A_C | EyeGex | Replicated | Non-retina; enhancer |
| TMEM97/VTN | TMEM199 | chr17_28376663_T_A | Orozco *et al.* | Replicated | |
| TNFRSF10A | TNFRSF10A | chr8_23231471_C_G | Orozco *et al.* | No METR eQTL | Non-retina; enhancer |
| TRPM1 | TRPM1 | chr15_31080689_T_C | Orozco *et al.* | METR-eQTL in high LD | Non-retina; enhancer |

***Over 800 eGenes are newly identified in the NSR and RPE***

To evaluate the tissue-specificity of our dataset, we compared the METR-eQTLs with non-eye specific eQTLs from the GTEx project ***(Figure 2C)***. We identified 337,424 METR-eQTLs (22.7%) and 916 eGenes (8.7%) that had not been previously identified by GTEx (***Figure 2C***); 251,685 (74.6%) of these eQTLs have not been previously described as eQTLs in NSR or RPE previously. Of the novel eGenes, 479 (57.9%) encoded lncRNAs and 5 had previously been associated with a known rare monogenic eye-disease (*HPS4, ACO2, CRX, CRYAA, PEX26*). We evaluated the degree of similarity between METR-eQTLs and eQTLs from each GTEx tissue using the Intersection over Union (IoU) statistic, which accounts for the wide variation in the number of eQTLs from different tissues (***Supplementary figures 9 and 10***). Brain cortex had the highest level of eQTL similarity to our dataset (IoU = 0.28), and 5 of the top 10 most similar tissues were from the brain.

***Genetic variants driving expression profile differences are enriched in candidate cis-regulatory elements (cCREs), with highest enrichment in retina-specific cCREs***

To understand whether eQTLs were enriched for putative regulatory regions, we compared locations of METR-eVariants to cell-type agnostic cis-candidate regulatory elements (cCRE) available through ENCODE (V3). METR-eVariants were enriched in cell-type agnostic promoters ($p= 8.05\times10^{-19}$) and proximal enhancers ($p = 8.48\times10^{-26}$), compared to control variants matched for allele frequency and gene density. There was no enrichment of eVariants in distal enhancers, CTCF binding sites or DNase-H3K4me3 sites ***(Figure 3A)***.

When stratified by cell-specific regulatory regions, bootstrapping analysis indicated a significant enrichment of METR-eVariants in NSR-specific (p-value = $4.52\times10^{-28}$) and RPE-specific cCREs (p-value = $8.74\times10^{-10}$) (Cherry *et al.*, 2020) compared to control variants matched for allele frequency and gene density (number of gene TSSs within 1Mb of variant) ***(Figure 3B).*** Furthermore, we observe a significant enrichment of METR-eVariants in cell-type specific accessible chromatin regions across 8 different retina cell types (Wang *et al.*, 2022), with the greatest enrichment in rod cells (p-value = $6.69\times10^{-58}$) and cone cells (p-value = $6.58\times10^{-57}$) ***(Figure 3C).*** Non-eye cCREs from adult tissues in EpiMap were also enriched for METR-eVariants, although the enrichment was lower than in the NSR and RPE. Despite the relative enrichment in annotated regulatory loci, most METR-eVariants (88.2%) do not overlap with any previously characterised cCREs ***(Figure 3D)***.

Further, we assessed whether eVariants previously implicated in AMD risk intersected with cell-type agnostic or cell-specific regulatory regions, observing overlap for 48% (10/21) of unique eVariants with characterised cis-regulatory elements, although only 1 eVariant overlapped with regions previously shown to be active in the retina (Table 1).

***Properties of METR-eQTLs differ between known monogenic disease genes and non-disease-related genes***

To understand if there were trends that were specific to eQTLs associated with known monogenic eye disease genes, we compared findings from this study against the EyeG2P resource (Lenassi *et al.*, 2023). We identified 230 METR-eGenes that were described as causes of rare monogenic disorders in EyeG2P (eye-disease genes) and compared trends identified in these genes against all other METR-eGenes (n = 10,241) (eye non-disease genes) (*Figure 4*). We observed significantly lower expression variability across samples for eye disease eGenes compared to eye non-disease eGenes ($p<2.2\times10^{-16}$) (***Figure 4A***). Eye disease eGenes were associated with significantly fewer eQTLs per gene ($p=7.8\times10^{-3}$) ***(Figure 4B)***. Additionally, eQTLs associated with eye disease eGenes have a significantly lower impact on gene expression ($p<2.2\times10^{-16}$) (***Figure 4C)***, and significantly higher allele frequency (AF, gnomAD v4) compared to eye non-disease eVariants ($p<2.2\times10^{-16}$) ***(Figure 4D***). Genes that have been associated with rare monogenic eye disease have higher expression (mean TPM = 37.9) than non-disease genes (mean TPM = 20.3), and to control for this potential confounding factor, we adopted a bootstrapping approach (n = 1000 iterations) to randomly resample 100 eQTLs associated with eye-disease genes and 100 eQTLs associated with non-eye disease genes matched for gene expression level (± 5% TPM) (***Supplementary figure 11***). The direction of trends remained similar after bootstrapping, with lower effect sizes and higher allele frequencies observed for eQTLs associated with eye disease genes than non-disease genes (***Supplementary figure 11***). For both eye disease eGenes and eye non-disease eGenes, there is a negative correlation ($p<2.2\times10^{-16}$) between eVariant allele frequency and the impact of each eQTL on gene expression ***(Figure 4E).*** These findings are consistent with the hypothesis that eVariants which are more common in the population have lower effect sizes on gene expression compared to rarer eVariants (min eVariant allele frequency = 2.5%), and are suggestive of a selective bias against rarer eVariants impacting known eye disease genes.

## Rare variants are plausible drivers of transcriptomic outliers in NSR and RPE

We utilised the DROP workflow (Yépez *et al.*, 2021) to identify statistical outlier events within the METR transcriptome datasets, including expression, splicing and allelic imbalance outliers ***(Table 2).*** We identified 1,051 unique instances of a gene being aberrantly expressed in an METR-sample (METR expression outlier events, METR-eOutlier events) (adjusted $p<0.05$); 728 of these events were in the NSR, 443 in the RPE, and 120 eOutlier events were found in both tissues. A median number of 3 genes per sample were considered significant outlier events in the NSR (IQR=1,4) and 1 in the RPE (IQR=0,2). In total, we tested 3,209,821 gene-sample events in the NSR and 3,050,081 in the RPE, indicating a significant outlier rate of 0.023% and 0.015%, respectively. These observations are consistent with a recent study of the GTEx cohort, describing significant outlier rates of 0.026% (Hölzlwimmer *et al.*, 2025).

*Table 2* ***Transcriptome outliers detected in the METR-cohort.*** *The DROP pipeline was utilised to identify three types of transcriptome outliers (expression, splicing and monoallelic expression) in the NSR and RPE.*

| **Type of outlier** | **Expression** | | | **Splicing** | | | **Allelic balance** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Tissue** | **NSR** | **RPE** | **Shared** | **NSR** | **RPE** | **Shared** | **NSR** | **RPE** | **Shared** |
| **Unique events (count)** | 728 | 443 | 120 | 8,610 | 8,923 | 469 | 37,230 | **106,116** | **9,888** |
| **Unique genes (count)** | 702 | 439 | 137 | 6,095 | 6,270 | 2,972 | 7,893 | **17,167** | **7,410** |
| **Number of genes identified as outliers, per sample (median)** | 3 | 1 | 0 | 12 | 6 | 1 | 173 | 138 | 53.5 |
| **IQR (Q1, Q2)** | (1, 4) | (0,2) | (0, 1) | (8,18) | (3, 24) | (0,2) | (146, 203) | (71, 1,003) | (34,87) |
| **Range (min, max)** | (0, 64) | (0, 46) | (0, 24) | (2, 1,948) | (0, 1,258) | (0, 209) | (102, 2,823) | (14, 4,308) | (8, 214) |

For each eOutlier event, we were able to harness paired genomic data to identify candidate rare variants potentially driving aberrant expression profiles. We leveraged a hierarchical framework and a probabilistic model to prioritise candidate rare genetic variants driving changes in expression. This identified 230 (23%) eOutlier events likely driven by protein-coding variants and 314 (31%) events with non-coding candidate variants (***Supplementary table 6***).

***Rare variants predicted to have a functional impact are identified for 50% of eOutlier events in NSR and RPE***

First, we applied a hierarchical framework to identify rare SVs, CNVs and SNVs which were predicted to result in loss-of-function (pLoF, including frameshift, nonsense and start/stop site loss variants) or were expected to disrupt a nearby non-coding regulatory region (***Supplementary figure 12***). Following this approach, we identified candidate functional variants driving 528 eOutlier events (50.2% of all eOutlier events identified in this study) *(**Figure 5A; Supplementary table 6**).* Of these, 131 eOutlier events were co-occurring with a SV or CNV impacting the coding-sequence of the outlier gene (77 NSR-only,12 RPE-only and 42 in both tissues), and 98 with a pLoF SNV (77 NSR-only, 6 RPE-only and 15 in both tissues) impacting the same gene (***Figure 5a***). For those eOutlier events not explained by an SV, CNV or pLoF SNV disrupting the coding sequence, we identified genomic variants in 71 eOutlier events that were within 10Kb of the gene body and impacted an eye-specific cCRE (Cherry *et al.*, 2020), including SV/CNVs (n=2) and SNVs that are rare (<0.01 AF) or absent in gnomAD (n=69). We also identified non-eye specific cCREs from EpiMap (Boix *et al.*, 2021) which were disrupted by SV/CNVs (n=23) or rare SNVs (n=121) within 10 Kb of the eOutlier gene. Examples of rare variants identified through this analysis strategy are included in ***Figure 6 and Supplementary figure 13.***

***A probabilistic model demonstrates high concordance for candidate SNVs driving expression outlier profiles in NSR and RPE***

Next, we applied Watershed (Ferraro *et al.*, 2020), a probabilistic model that was re-trained on 6 tissue-specific outlier *p*-values from DROP. This was used in the METR transcriptome

datasets to obtain posterior probabilities for SNVs and small indels that may be driving outlier expression profiles. Eye-specific cCREs were added as annotation features for Watershed (***Supplementary Table 7***), and identified 135 (13%) eOutlier events that were likely to be caused by nearby rare variants (posterior probability > 0.8), of which 110 (81%) were also predicted to be driven by the same variants by the hierarchical model and 11 were predicted to be driven by SV/CNVs that are not considered by Watershed (***Figure 5B****).* We used bootstrapping analysis to compare the annotations associated with these variants to other rare variants which overlapped with eOutlier genes but were not predicted by Watershed to have a functional impact, observing an enrichment of canonical splice variants, frameshift variants, stop gain variants and variants predicted to disrupt splicing ***(Figure 5C)***. In support of other analyses described in this study, there was an enrichment of rare variants which overlap with retina cCREs, and a slight enrichment of rare variants which overlap with epigenomic marks associated with non-eye specific regulatory elements from ENCODE. In total, there were 34 eOutlier events where the functional variants prioritised by Watershed intersected with a known candidate cis-regulatory region (cCRE); 28 of these cCREs were active in the eye.

### ***Functional assays confirm the impact of rare variants prioritised as drivers of eOutlier expression***

A dual reporter luciferase assay was performed in Human K562 cells to investigate the impact of a *CAND2* heterozygous variant that was prioritised as a driver of drastically reduced expressed in NSR (fold change = 0.6, Z-score = -5.6, p-adj = 0.004) and RPE (fold change = 0.5, Z-score =-4.7, p-adj = 0.049) due to its overlap with features indicative of the *CAND2* promoter region (NM_001162499.2:c.-41A>G; Figure 7). *CAND2* has recently been implicated in AMD risk through GWAS meta-analysis in European ancestry populations (Gorman *et al.*, 2024) and has an emerging role in the targeted degradation of proteins that is distinct from its *CAND1* homolog (Wang *et al.*, 2025). Our dual reporter luciferase assay reveals a significant reduction in *CAND2* promoter activity in the presence of c.-41A>G (adj p=0.005; Figure 7), confirms the disruption of *CAND2* activity in NSR and RPE, and provides a proof-of-principle for our applied prioritisation methods for genetic drivers of expression outliers.

## Discussion

We present a unique resource to interrogate the impact of both common and rare genomic variation on gene regulation in the human NSR and RPE. We characterised novel eQTL associations that are tissue-specific (***Figure 2***) and are enriched to known promoters and proximal enhancers (***Figure 3***). We show that eQTLs impacting genes known as a cause of rare genetic eye disease have different properties when compared to those genes which are not known as a cause of eye disease (***Figure 4***). We also identify candidate non-coding rare variants, SVs and CNVs which impact cCREs and represent plausible drivers of outlier expression profiles in human NSR and RPE (***Figure 5***), including functional validation of a prioritised non-coding genetic variant impacting *CAND2* (***Figure 7***). The METR resource can be used alongside other multi-omic datasets to facilitate discovery of novel eye-specific regulatory elements, including those implicated in common (e.g. AMD) and rare (e.g. IRDs) genetic disorders impacting the retina.

The cohort of 201 human donors described in this study represents the first dataset, to our knowledge, to pair whole genome sequencing with high-depth RNA sequencing data from the NSR and RPE. Previous studies have developed RNA sequencing from the NSR alongside genotyping arrays (Ratnapriya *et al.*, 2019; Orozco *et al.*, 2020; Strunz *et al.*, 2020) and this has enabled the characterisation of eQTLs in the retina including preliminary data supporting the role of a limited number of eQTLs in AMD (***Table 1***). We performed extensive QC for both DNA and RNA sequencing data to confirm the validity of the datasets generated in this study, in particular due to the prolonged median ischemic time compared to GTEx samples which may impact quality of the data obtained (Ferreira *et al.*, 2018). These analyses confirmed suitable RNA integrity values across the cohort (***Supplementary figure 1***), high unique mapping rates with appropriate read lengths and appropriate 3'/5' biases from RNAseq data (***Supplementary figures 1 & 4***), along with representative gene expression profiles (***Figure 1C, Supplementary Table 4***). The high-depth and high-quality RNA and whole genome sequencing datasets developed for this study are from a cohort of individuals without clear signs of late-stage AMD and have enabled novel biological insights beyond those described previously.

Firstly, we were able to assess whether previously characterised eQTLs are replicated utilising alternative methods and technologies in an independent cohort of individuals of European genetic ancestry without late-stage AMD (***Supplementary figure 3***). As gene expression profiles have been shown to be significantly disrupted during AMD pathogenesis (Voigt *et al.*, 2022; Orozco *et al.*, 2023), it is important to identify eQTL signals that are amplified or disrupted by broader changes in transcriptome profiles associated with AMD, as well as those that remain consistent within a cohort of individuals without clear signs of late-stage AMD. Overall, we show high levels of replication of eQTL findings from Ratnapriya et al, with 5,993 identical eGenes and replication of 13 eQTLs previously implicated with a role in AMD (***Table 1***), including *PILRB* which has recently been shown to lead to photoreceptor dysfunction in mice when function is impaired (Dey *et al.*, 2025). Notably, we identified 5 novel QTLs for genes previously implicated in AMD (*ACAD10, HTRA1, B3GLCT,*

*PLA2G12A, BAIAP2L2*) and 4 genes implicated in AMD without replication of a previously characterised eQTL (*CFI, COL4A3, RDH5, TNFRSF10A*). This suggests that differences in the approaches undertaken and/or cohort composition, e.g. AMD status, cohort size and/or genetic ancestry, impacts the influence of genomic drivers on expression of these genes.

Second, variants which are rare in the population or unique to individuals have been demonstrated to drive drastic changes in expression profiles, so called 'expression outliers', across different tissues (Li *et al.*, 2017; Ferraro *et al.*, 2020). The use of complete genomic sequencing in this cohort, achieving a median coverage of 36x, has enabled the characterisation of a greater diversity of genomic variation than has previously been studied in the context of expression drivers in the NSR and RPE, and identified thousands of new regions which can be interrogated for rare variation within disease cohorts (Ellingford *et al.*, 2022). Using two distinct variant prioritisation approaches, we describe rare variants in the general population, including SVs, CNVs and small variants that are the most likely drivers of expression outliers in these tissues (***Figure 5***). Through functional validation of a prioritised non-coding variant in the *CAND2* promoter region, we establish a proof-of-principle for the applied variant prioritisation approaches (***Figure 7***), and provide mechanistic insight of non-coding regions regulating the expression level of AMD-risk associated genes (Gorman *et al.*, 2024). These data encourage further functional follow-up for the 578 prioritised variants that may be causative of pronounced changes in expression profiles in the human retina, including 272 rare variants predicted to cause loss-of-function and 299 that intersect with non-coding regions (including examples presented in ***Figure 6*** and ***Supplementary figure 13***). Other recent studies have identified outlier-associated non-coding rare variants that contribute to common disease predisposition (Smail *et al.*, 2022) and underpin rare genetic disorders (Wakeling *et al.*, 2022; Tenney *et al.*, 2023). Moreover, non-coding variation has been identified as a cause of genetic ophthalmic disorders in untranslated regions (Dueñas Rey *et al.*, 2024), retina-specific exons (Vig *et al.*, 2020), promoters (Daich Varela *et al.*, 2023), distal enhancers (Small *et al.*, 2016) and non-coding genes (Quinodoz *et al.*, 2025) expressed in the NSR and RPE. With the increasing availability of genomic sequencing datasets for the diagnosis and discovery of genetic disorders (Turnbull *et al.*, 2018; The 100,000 Genomes Project Pilot Investigators, 2021), including ophthalmic conditions (Ellingford *et al.*, 2016), the data developed in this study is timely and provides an opportunity, alongside other complementary datasets (D'haene *et al.*, 2024), to identify new pathogenic mechanisms underpinning genetic disorders.

Third, we have generated high-coverage RNAseq datasets achieving, on average, 139 million uniquely mapping reads for NSR and 62 million uniquely mapping reads for RPE. Previous studies have developed lower coverage RNAseq datasets for NSR, for example, EyeGEx (Ratnapriya *et al.*, 2019), Orozco *et al* (Orozco *et al.*, 2020) and Pinelli *et al* (Pinelli *et al.*, 2016) generated 33, 30 and 72 million sequencing reads per sample, respectively. Previous studies have remarked on the level of transcript diversity in NSR (Farkas *et al.*, 2013), and highlighted the advantage of high-depth RNAseq in this context. In comparison to EyeGEx, our high coverage approach elevates the number of observable protein-coding

genes by 23% (from 13,662 to 16,765) and newly identifies 3,663 eGenes. For 4,481 eGenes that are not replicated from EyeGEx, further study, for example harmonisation of genomic and RNA sequencing dataset processing and meta-analyses would assist in understanding whether their detection is influenced by cohort composition, methodologies undertaken and/or sample size. The increased number of eGenes from this study enables observation of patterns in gene expression at increased resolution and has granted insight into the trends associated with genes previously implicated in genetic disorders impacting vision. Overall, we observed that eGenes that have been characterised as a cause of rare genetic eye disease (Lenassi *et al.*, 2023) have lower expression variability across individuals than non-disease genes (***Figure 4***), suggesting that regulation of these genes is more tightly controlled in NSR and RPE. The role of eQTLs in genetic disorders remains incompletely understood. For example, whilst some studies have shown eQTLs contribute to onset, penetrance and expressivity (Castel *et al.*, 2018; Einson *et al.*, 2023), including genetic disorders impacting the eye (Michaud *et al.*, 2022), others have found limited evidence for their role in neuronal genetic disorders (Rio Frio *et al.*, 2008; Wigdor *et al.*, 2024). Here, we observe that eQTL variants which were associated with changes in expression of eye-disease genes had significantly lower effect sizes and their allele frequency was higher than eQTLs impacting genes that have not previously been implicated in eye disease (***Figure 4; Supplementary figure 11***). Intuitively, the absence of rarer and higher impact eVariants amongst a population of individuals without signs of genetic eye disease suggests constraint on genomic variation with these properties, although population-scale modelling and statistical analysis is required to formally test this hypothesis.

Finally, as our cohort includes 158 individuals with RNA extracted and sequenced from both NSR and RPE datasets, this enables further insights into the expression patterns and regulatory architecture of these tissues, unbiased by sample preparation methods and/or differences between individuals, e.g. genomic background. It should be noted that our cohort is biased towards male individuals (64%) and this may have a hidden bias on eQTLs and transcriptome differences identified. However, our data newly identifies 916 eGenes in NSR and RPE compared to those characterised in other tissues (THE GTEX CONSORTIUM, 2020), and we observe high level of overlap in eGenes between NSR and RPE, including 86% of RPE eGenes and 32% of NSR eGenes. These data further support the high level of overlap previously observed for active enhancers and promoters between RPE/choroid and NSR (Cherry *et al.*, 2020).

Whilst the findings of this study have enhanced our understanding of genomic regulation in human NSR and RPE, other approaches that utilise single-cell (Lukowski *et al.*, 2019; Menon *et al.*, 2019; Orozco *et al.*, 2020; Yan *et al.*, 2020; van Zyl *et al.*, 2022) single-nuclei (Liang *et al.*, 2019; Monavarfeshani *et al.*, 2023) and spatial (Choi *et al.*, 2023; Dorgau *et al.*, 2024) transcriptomic approaches enable increased precision to understand gene expression in specialised retinal layers and cell types. These approaches are particularly advantageous for the NSR, which is a highly heterogeneous tissue comprised of several specialised layers and neuronal cell types, including photoreceptors, bipolar cells, amacrine cells and

horizontal cells (Masland, 2012), and where transcriptome profiles may differ substantially between the central and peripheral retina (Sharon *et al.*, 2002) To overcome potential shortcomings of the bulk RNAseq approach adopted in this study, we performed deconvolution analyses to estimate the relative sample composition against single-nuclei RNA-sequencing (Monavarfeshani *et al.*, 2023). Given the complexity associated with retinal tissue dissection and storage (McHarg *et al.*, 2015; Cabral *et al.*, 2017), the deconvolution approach also enabled confirmation of tissue sample integrity alongside differential expression profiles (***Figure 1C***). Bulk RNAseq from NSR had representation, as expected, from diverse cell types with significant enrichment towards rod photoreceptors and retinal astrocytes, representing >50% of the estimated cellular make-up of most samples. In keeping with current understanding of retinal ageing (Gao and Hollyfield, 1992) there is observed significant loss of rod photoreceptors with age (***Supplementary Figure 5B***). However, deconvolution is naturally limited by the relative differential transcriptional activity between cell types, and is complicated by cell types with similar transcription profiles, for example, between Müller glia and retinal astrocytes (Yan *et al.*, 2020). We expect that the high number of astrocytes predicted in RNA samples is influenced by similar transcription profiles to other cell types, and whilst we confirm that we have generated high-quality RNA sequencing datasets (***Supplementary figures 1 & 4, Figure 1***), these estimates may also be influenced by altered transcriptome profiles in samples with longer ischemic times (Ferreira *et al.*, 2018) and/or the high sequencing depth coverage generated. Moreover, the retina is known to have cyclic patterns of gene expression, related to both circadian rhythm and natural function, i.e. response to light (Bhoi *et al.*, 2023), and as such there is incomplete molecular understanding of all cell types present in the human retina (Monavarfeshani *et al.*, 2023). Overall, these data support the integrity of the RNAseq dataset developed in this study and whilst confident quantification of the cell types present is not possible, our analyses confirm that the datasets are representative of major cell types in the retina.

Taken together, the data presented in this study provide new insights into the genomic control of gene regulation in the human retina. We build upon previous understanding through replication of eQTLs in a cohort of individuals without clear signs of late-stage AMD, characterise hundreds of new genes under genomic regulation, and provide insights into the role of rare variants, SVs and CNVs in disruption of gene expression in these specialised tissues that enable vision. Future studies utilising this resource, including meta-analysis with other published datasets, co-localisation and transcriptome-wide association studies incorporating findings from genome-wide association studies, will continue to develop understanding of the expression profiles and the role of non-coding genetic variation in the onset and presentation of genetic disorders impacting vision.

## Acknowledgments

We express our sincere thanks to the donors, and their families, for enabling this research. This research was supported by the Macular Society (United Kingdom), Fight For Sight, the UK Medical Research Council and the NIHR Manchester Biomedical Research Centre (NIHR203308). This work has also been supported by the Wellcome Trust (224643/Z/21/Z, Clinical Research Career Development Fellowship, P.I.S), the UK National Institute for Health Research (NIHR) Clinical Lecturer Programme (CL-201-06-001, P.I.S), the NIHR Research Professorship grant (RP-2016-07-011, DB) and the NIH/NEI (R01 EY031424 to A.V.S and R01EY035717 to K.M.B.). We thank Selina Mcharg, Nadhim Bayatti and Jay Brown for the development of the Manchester Eye Tissue Resource. We also thank staff at the University of Manchester Genomic Technologies Core Facility, and at the Ocular Genomics Institute, Harvard Medical School, for their help in the generation of DNA and RNA sequencing datasets for this study. The views expressed are those of the authors and not necessarily those of the funders, including the NIHR and the Department of Health and Social Care.

## Data Availability

All raw RNA sequencing and genomic sequencing datasets generated in this study have been made available through the European-Genome phenome Archive (EGA; Study ID: EGAS50000001443; Dataset: EGAD50000002082). Processed datasets, including eQTL results, eOutlier statistics and aggregated genomic variant files are also available through EGA.

# Methods

## Gene expression quantification in Neurosensory Retina and Retinal Pigment Epithelium from RNA-Seq data

Paired-end short-read sequencing of polyA-enriched mRNA (RNAseq) was performed on an Illumina NovaSeq 6000 instrument for two layers of the retina: (1) the entire neurosensory retina (NSR), including macula and peripheral regions, and (2) pelleted cells from the retinal pigment epithelium (RPE), which were scraped from Bruch's membrane. Donor eye tissues were obtained from the Manchester Eye Tissue Repository, an ethically approved Research Tissue Bank (UK NHS Health Research Authority, 15/NW/0932). Eye tissue was acquired after the corneas had been removed for transplantation and explicit consent had been obtained to use the remaining tissue for research. Samples were selected for RNAseq with reference to RNA concentration (ng / µl) and integrity (RNA Integrity Number – RIN) values, calculated with the Agilent TapeStation system. (*see Supp Methods 1.1 and 1.2 for additional details on tissue extraction and the RNA sequencing protocol*).

The Genotype-Tissue Expression (GTEx) analysis pipeline (GTEx Consortium, 2019) was applied to RNAseq datasets to assess quality, and to perform alignment and expression quantification. Alignment was performed against the GRCh38 human reference genome using STAR v2.7.4a (Dobin *et al.*, 2013). Duplicate reads were marked with Picard v2.27.1 ('Picard toolkit', 2019). Gene-level expression quantification, using the GENCODE v38 annotation (Frankish *et al.*, 2019) was carried out using RNA-SeQC 1.1.9 (DeLuca *et al.*, 2012), for gene-level read counts and RSEM v1.3.0 (Li and Dewey, 2011), for gene-level quantifications in transcripts per million (TPM). Quality assessments of processed RNAseq datasets included reference to the total number of reads, number of uniquely mapped reads, number of splice junctions, number of chimeric reads, read length and 3'/5' bias for all NSR and RPE samples. To ensure concordance between paired WGS and RNAseq samples, we excluded WGS-RNAseq pairs where the predicted relatedness, calculated using *Somalier* (Pedersen *et al.*, 2020), was <0.8.

## Whole Genome Sequencing Data

Short-read paired-end whole genome sequencing (WGS) was generated for each donor on an Illumina NovaSeq6000 Instrument using DNA extracted from iris biopsies *(see Supp Methods section 2 for additional details)*. Genome alignment and variant calling was carried out using Illumina DRAGEN 4.0.3 software with Machine Learning and Graph Map Enabled. Aggregate variant detection and harmonisation was carried out using Illumina DRAGEN 4.0.3 software Population Mode. We applied quality control filters to the aggregate VCF to remove low-quality variant calls using a combination of *bcftools* (v.1.16) and *PLINK* (v.2.0) (*see Supp Methods 2.3).* For eQTL mapping, aggregate genotypes were binarized using PLINK 2.0.

## Cell type deconvolution of bulk RNA-seq data

We used BayesPrism (Bayesian cell Proportion Reconstruction Inferred using Statistical Marginalization) (Chu *et al.*, 2022) to run a deconvolution model to estimate the proportion of retinal cell types in the generated bulk RNA-seq data in NSR and RPE. The reference dataset to train the model was a single-cell RNAseq dataset from the ocular posterior segment (Monavarfeshani *et al.*, 2023) *(See Supp Methods 1.5 for additional details).*

**Differential expression analysis between NSR and RPE**

To ensure the validity of the transcriptomic datasets generated in this study, we assessed the biological relevance of expressed genes in NSR and RPE. We used the R package *deseq2* (Love, Huber and Anders, 2014) to identify genes that were differentially expressed between NSR and RPE. We included age and sex as covariates in the *deseq2* model with the false discovery rate threshold set at 0.05. To confirm *deseq2* results, we replicated the differential expression analysis using *edgeR* (Robinson, McCarthy and Smyth, 2010).

To identify which gene ontology biological pathways were enriched in the upregulated genes in NSR/RPE, we carried out gene set enrichment analysis (GSEA) of the genes which were differentially expressed between both tissues (FDR < 0.05), using *WebGestalt* (Liao *et al.*, 2019). We processed the GSEA output with a clustering algorithm, *rrvgo* (Sayols, 2023), to group similar GO terms together and selected representative terms. *(See Supp Methods 1.6 for additional details).*

**Input Data for cis-eQTL analysis**

For eQTL analysis, we generated a normalised expression matrix for each tissue. Genes which did not meet expression thresholds of >0.1 TPM in at least 20% of samples and ≥6 reads in at least 20% of samples were removed from eQTL analysis. Expression values were normalised using the trimmed mean of M-values normalisation (TMM) method (Robinson and Oshlack, 2010) and using an inverse normal transform.

To account for known and unknown biological and experimental confounding factors, a set of 30 covariates were generated for each RNA-Seq sample using the Probabilistic Estimation of Expression Residuals (PEER) method (Stegle *et al.*, 2010) applied to normalised gene expression levels.

Principal component analysis with EIGENSOFT 6.0.1 (Patterson, Price and Reich, 2006) was carried out to capture ancestral variation within the cohort. The top five principal components for each participant were used as covariates in the eQTL analysis.

**Cis-eQTL mapping with tensorQTL**

TensorQTL (Taylor-Weiner *et al.*, 2019) was used to identify genetic variants which were significantly associated with the expression of nearby genes (up to 1 Mb away) in NSR and RPE (FDR < 0.05). The required input files were the normalised gene expression matrix, the binary and filtered genotype data and a covariates table which included the following

information for each participant: sex, WGS batch, five top principal components and 30 PEER factors. To quantify the eQTL effect size we estimated the log2 allelic fold change (aFC), following the method established by Mohammadi et al. (2017) (*see Supp Methods section 4 for additional details)*

**Comparison with other eQTL studies**

We compared all METR eQTLs with retina eQTLs mapped by EyeGEx (Ratnapriya *et al.*, 2019) and (Strunz *et al.*, 2020). We identified genes had been associated with eQTLs in our study and in EyeGEx and/or Strunz et al. (2020) (eGenes shared between studies). For these shared eGenes, we extracted the top eQTLs identified by EyeGEx and/or Strunz et al. (2020) and checked if they were replicated in our cohort or if they were in high LD ($r2 > 0.8$) with a METR-NSR eQTL. Pairwise LD scores were calculated using LDlinkR (Myers, Chanock and Machiela, 2020).

To compare to non-retina tissues, all significant eQTL associations were downloaded from the GTEx Open Access portal (v8) for each available tissue (https://www.gtexportal.org/home/downloads/adult-gtex/qtl). We calculated the intersection between the number of METR-eQTLs and eGenes which were also shared by each GTEx tissue using the Intersection over Union (IoU) statistic. The IoU calculates the ratio of the number of eQTLs/eGenes present in both sets over the total number of eQTLs/eGenes in one set and/or the other.

**Annotation of eVariants and bootstrapping analysis to calculate enrichment of eQTLs in characterised regulatory loci**

All NSR and RPE eQTL variants were annotated with the Ensembl Variant Effect Predictor (McLaren *et al.*, 2016). We assessed overlap and annotated all eVariants with a set of tissue-specific and cell-type specific annotations of candidate cis-regulatory elements (cCREs) from a variety of sources (*Supp Methods Table 1*). These included characterised regulatory loci from retina, RPE and macula (Cherry et al. (2020), cell-type specific regions of open chromatin detected by scATACseq from retina samples (Wang *et al.*, 2022), non-eye specific cCREs from adult tissues in EpiMap (Boix *et al.*, 2021) and cell-type agnostic candidate cis-regulatory element (cCRE) annotations from ENCODE (Snyder *et al.*, 2020).

To calculate the relative enrichment of eVariants which overlapped with each type of regulatory element, we used bootstrapping analysis. We carried out 1,000 iterations by subsampling 100,000 random eVariants with replacement and 100,000 control variants from our cohort that were included as input for the eQTL mapping and did not meet the eQTL significance threshold (FDR < 0.05), matched for gene density and allele frequency. We compared the ratio of eVariants to control variants that intersected with each type of regulatory element. (see *Supp Methods section 5 for additional details)*

**Analysis of the properties of eQTLs that impact known eye disease-related genes**

To understand if there were trends that were specific to eQTLs associated with known monogenic eye disease genes, we utilised the EyeG2P resource (eye-disease genes) (Lenassi *et al.*, 2023). All other METR-eGenes were considered non-eye disease genes. We compared a eQTL/eGene properties between eye-disease and non-eye disease gene eQTLs, including eVariant allele frequency and effect size, measured using log2 allele fold change. *(See Supp Methods section 6 for additional details).*

**Identification of transcriptome outliers using the DROP pipeline**

We utilised the DROP v.1.4.0 pipeline (Yépez *et al.*, 2021) to identify transcriptome outliers from NSR and RPE, using standard parameters.

**Hierarchical workflow to identify candidate variants driving outlier expression**

We developed a hierarchical workflow to identify candidate variants driving outlier expression (eOutliers) using snakemake version 7.32 (Supp *Methods Figure 1)*. Briefly, the workflow would first search for a pLoF variant from the eOutlier sample in the corresponding eOutlier gene, which could be an exonic structural variant, or a SNV with a high impact consequence based on *Ensembl's Variant Effect Predictor (v.112.0).* If no pLoF variant could be identified, the workflow would then search for regulatory variants which were within 10Kb of the eOutlier gene body. Regulatory variants were defined as structural variants and rare SNVs (gnomAD AF < 0.01) which overlapped with nearby retina cCREs or non-retina specific cCREs from different adult tissues in EpiMap. If no regulatory variant was identified, the model would identify any other non-coding structural variant within 10Kb of the eOutlier gene body, before returning a negative search result (*see Supp Methods section 8 for additional details).*

**Implementation of Watershed**

For all genes with an eOutlier in the NSR, we extracted all rare variants (gnomAD allele frequency < 1%) which intersected with the gene body ± 10Kb. Variants were extracted for all samples with NSR RNAseq data from the post-QC aggregate VCF. We annotated all rare variants with selected annotations from VEP (McLaren et al., 2016) and CADD (Schubach et al., 2024) (Supplementary Table 6) and intersected them with known retina-specific cCREs from Cherry et al. (2020) and non-retina specific cCREs from EpiMap. Missing annotations were replaced with default imputation values obtained from CADD (Supplementary Table 6). The Watershed model (Ferraro *et al.*, 2020)was run using the predict_watershed.R script with an adjusted *p*-val threshold of 0.05 and the number of dimensions set to 6. (*See Supp Methods section 9 for additional details).*

**Dual reporter luciferase assay**

A 294bp fragment of the wild-type promoter region from *CAND2* was PCR-amplified from control genomic DNA using Phusion High-Fidelity DNA Polymerase (Promega). To introduce variants, two overlapping fragments were amplified using combination of mutagenic primers. Variants constructed were the variant of interest, NM_001162499.2:c.-41A>G, and a variant that is common in the general population and not expected to impact *CAND2* expression, NM_001162499.2:c.-36C>T.

The wild-type and variant fragments were assembled into NheI-NcoI digested pGL4.10[luc2] firefly luciferase plasmid using the Gibson method. The assemblies were transformed into competent E. coli grown overnight on LB agar containing carbenicillin. Candidate colonies were picked for culture and plasmid isolation. The plasmid constructs were verified by Sanger sequencing. Human K562 cells were transiently transfected with 500 ng of plasmid using Lipofectamine LTX (Invitrogen) following the manufacturer's standard protocol. Empty pGL4.10[luc2] plasmid was transfected as a control for background activity. The Renilla luciferase pGL4.74[hRluc/TK] vector (Promega) was co-transfected as an internal luminescence control. Following 20-24 hr incubation at 37°C with 5% $CO_2$, a dual luciferase assay was conducted using the Dual-Glo® Luciferase Assay (Promega).

## Bibliography


Barbeira, A.N. *et al.* (2021) ‘Exploiting the GTEx resources to decipher the mechanisms at GWAS loci’, *Genome Biology*, 22(1), p. 49. Available at: https://doi.org/10.1186/s13059-020-02252-4.

Bhoi, J.D. *et al.* (2023) ‘Circadian clock organization in the retina: From clock components to rod and cone pathways and visual function’, *Progress in Retinal and Eye Research*, 94, p. 101119. Available at: https://doi.org/10.1016/j.preteyeres.2022.101119.

Boix, C.A. *et al.* (2021) ‘Regulatory genomic circuitry of human disease loci by integrative epigenomics’, *Nature*, 590(7845), pp. 300–307. Available at: https://doi.org/10.1038/s41586-020-03145-z.

Cabral, T. *et al.* (2017) ‘Dissection of Human Retina and RPE-Choroid for Proteomic Analysis’, *Journal of Visualized Experiments: JoVE*, (129), p. 56203. Available at: https://doi.org/10.3791/56203.

Castel, S.E. *et al.* (2018) ‘Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk’, *Nature Genetics*, 50(9), pp. 1327–1334. Available at: https://doi.org/10.1038/s41588-018-0192-y.

Cherry, T.J. *et al.* (2020) ‘Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease’, *Proceedings of the National Academy of Sciences*, 117(16), pp. 9001–9012. Available at: https://doi.org/10.1073/pnas.1922501117.

Choi, J. *et al.* (2023) ‘Spatial organization of the mouse retina at single cell resolution by MERFISH’, *Nature Communications*, 14(1), p. 4929. Available at: https://doi.org/10.1038/s41467-023-40674-3.

Chu, T. *et al.* (2022) ‘Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology’, *Nature Cancer*, 3(4), pp. 505–517. Available at: https://doi.org/10.1038/s43018-022-00356-3.

Daich Varela, M. *et al.* (2023) ‘Multidisciplinary team directed analysis of whole genome sequencing reveals pathogenic non-coding variants in molecularly undiagnosed inherited retinal dystrophies’, *Human Molecular Genetics*, 32(4), pp. 595–607. Available at: https://doi.org/10.1093/hmg/ddac227.

Davenport, E.E. *et al.* (2018) ‘Discovering in vivo cytokine-eQTL interactions from a lupus clinical trial’, *Genome Biology*, 19(1), p. 168. Available at: https://doi.org/10.1186/s13059-018-1560-8.

DeLuca, D.S. *et al.* (2012) ‘RNA-SeQC: RNA-seq metrics for quality control and process optimization’, *Bioinformatics*, 28(11), pp. 1530–1532. Available at: https://doi.org/10.1093/bioinformatics/bts196.

Dey, P.N. *et al.* (2025) ‘Loss of paired immunoglobin-like type 2 receptor B gene associated with age-related macular degeneration impairs photoreceptor function in mouse retina’, *Human Molecular Genetics*, 34(1), pp. 64–76. Available at: https://doi.org/10.1093/hmg/ddae161.

D’haene, E. *et al.* (2024) ‘Comparative 3D genome analysis between neural retina and retinal pigment epithelium reveals differential cis-regulatory interactions at retinal disease loci’, *Genome Biology*, 25(1), p. 123. Available at: https://doi.org/10.1186/s13059-024-03250-6.

Dobin, A. *et al.* (2013) ‘STAR: ultrafast universal RNA-seq aligner’, *Bioinformatics*, 29(1), pp. 15–21. Available at: https://doi.org/10.1093/bioinformatics/bts635.

Dorgau, B. *et al.* (2024) ‘Deciphering the spatiotemporal transcriptional and chromatin accessibility of human retinal organoid development at the single-cell level’, *iScience*, 27(4), p. 109397. Available at: https://doi.org/10.1016/j.isci.2024.109397.

Dueñas Rey, A. *et al.* (2024) ‘Combining a prioritization strategy and functional studies nominates 5’UTR variants underlying inherited retinal disease’, *Genome Medicine*, 16(1), p. 7. Available at: https://doi.org/10.1186/s13073-023-01277-1.

Einson, J. *et al.* (2023) ‘Genetic control of mRNA splicing as a potential mechanism for incomplete penetrance of rare coding variants’, *Genetics*, 224(4), p. iyad115. Available at: https://doi.org/10.1093/genetics/iyad115.

Ellingford, J.M. *et al.* (2016) ‘Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease’, *Ophthalmology*, 123(5), pp. 1143–1150. Available at: https://doi.org/10.1016/j.ophtha.2016.01.009.

Ellingford, J.M. *et al.* (2022) ‘Recommendations for clinical interpretation of variants found in non-coding regions of the genome’, *Genome Medicine*, 14(1), p. 73. Available at: https://doi.org/10.1186/s13073-022-01073-3.

Farkas, M.H. *et al.* (2013) ‘Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes’, *BMC genomics*, 14, p. 486. Available at: https://doi.org/10.1186/1471-2164-14-486.

Ferraro, N.M. *et al.* (2020) ‘Transcriptomic signatures across human tissues identify functional rare genetic variation’, *Science (New York, N.Y.)*, 369(6509), p. eaaz5900. Available at: https://doi.org/10.1126/science.aaz5900.

Ferreira, P.G. *et al.* (2018) ‘The effects of death and post-mortem cold ischemia on human tissue transcriptomes’, *Nature Communications*, 9(1), p. 490. Available at: https://doi.org/10.1038/s41467-017-02772-x.

Fleckenstein, M. *et al.* (2021) ‘Age-related macular degeneration’, *Nature Reviews. Disease Primers*, 7(1), p. 31. Available at: https://doi.org/10.1038/s41572-021-00265-2.

Frankish, A. *et al.* (2019) ‘GENCODE reference annotation for the human and mouse genomes’, *Nucleic Acids Research*, 47(D1), pp. D766–D773. Available at: https://doi.org/10.1093/nar/gky955.

Fritsche, L.G. *et al.* (2016) ‘A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants’, *Nature Genetics*, 48(2), pp. 134–143. Available at: https://doi.org/10.1038/ng.3448.

Gao, H. and Hollyfield, J.G. (1992) ‘Aging of the human retina. Differential loss of neurons and retinal pigment epithelial cells’, *Investigative Ophthalmology & Visual Science*, 33(1), pp. 1–17.

Gorman, B.R. *et al.* (2024) ‘Genome-wide association analyses identify distinct genetic architectures for age-related macular degeneration across ancestries’, *Nature Genetics*, 56(12), pp. 2659–2671. Available at: https://doi.org/10.1038/s41588-024-01764-0.

GTEx Consortium (2019) ‘Laboratory and Analysis Methods’. Broad Institute of MIT and Harvard. Available at: https://gtexportal.org/home/methods (Accessed: 30 May 2023).

Hamel, A.R. *et al.* (2024) ‘Integrating genetic regulation and single-cell expression with GWAS prioritizes causal genes and cell types for glaucoma’, *Nature Communications*, 15(1), p. 396. Available at: https://doi.org/10.1038/s41467-023-44380-y.

Hanany, M., Rivolta, C. and Sharon, D. (2020) ‘Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases’, *Proceedings of the National Academy of Sciences of the United States of America*, 117(5), pp. 2710–2716. Available at: https://doi.org/10.1073/pnas.1913179117.

Hölzlwimmer, F.R. *et al.* (2025) ‘Aberrant gene expression prediction across human tissues’, *Nature Communications*, 16(1), p. 3061. Available at: https://doi.org/10.1038/s41467-025-58210-w.

Hoon, M. *et al.* (2014) ‘Functional architecture of the retina: development and disease’, *Progress in Retinal and Eye Research*, 42, pp. 44–84. Available at: https://doi.org/10.1016/j.preteyeres.2014.06.003.

Lenassi, E. *et al.* (2023) ‘EyeG2P: an automated variant filtering approach improves efficiency of diagnostic genomic testing for inherited ophthalmic disorders’, *Journal of Medical Genetics*, 60(8), pp. 810–818. Available at: https://doi.org/10.1136/jmg-2022-108618.

Li, B. and Dewey, C.N. (2011) ‘RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome’, *BMC Bioinformatics*, 12(1), p. 323. Available at: https://doi.org/10.1186/1471-2105-12-323.

Li, X. *et al.* (2017) 'The impact of rare variation on gene expression across tissues', *Nature*, 550(7675), pp. 239–243. Available at: https://doi.org/10.1038/nature24267.

Liang, Q. *et al.* (2019) 'Single-nuclei RNA-seq on human retinal tissue provides improved transcriptome profiling', *Nature Communications*, 10(1), p. 5743. Available at: https://doi.org/10.1038/s41467-019-12917-9.

Liao, Y. *et al.* (2019) 'WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs', *Nucleic Acids Research*, 47(W1), pp. W199–W205. Available at: https://doi.org/10.1093/nar/gkz401.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. Available at: https://doi.org/10.1186/s13059-014-0550-8.

Lukowski, S.W. *et al.* (2019) 'A single-cell transcriptome atlas of the adult human retina', *The EMBO journal*, 38(18), p. e100811. Available at: https://doi.org/10.15252/embj.2018100811.

Masland, R.H. (2012) 'The neuronal organization of the retina', *Neuron*, 76(2), pp. 266–280. Available at: https://doi.org/10.1016/j.neuron.2012.10.002.

McHarg, S. *et al.* (2015) 'Enrichment of Bruch's Membrane from Human Donor Eyes', *Journal of Visualized Experiments: JoVE*, (105), p. 53382. Available at: https://doi.org/10.3791/53382.

McLaren, W. *et al.* (2016) 'The Ensembl Variant Effect Predictor', *Genome Biology*, 17(1), p. 122. Available at: https://doi.org/10.1186/s13059-016-0974-4.

Menon, M. *et al.* (2019) 'Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration', *Nature Communications*, 10(1), p. 4902. Available at: https://doi.org/10.1038/s41467-019-12780-8.

Michaud, V. *et al.* (2022) 'The contribution of common regulatory and protein-coding TYR variants to the genetic architecture of albinism', *Nature Communications*, 13(1), p. 3939. Available at: https://doi.org/10.1038/s41467-022-31392-3.

Mohammadi, P. *et al.* (2017) 'Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change', *Genome Research*, 27(11), pp. 1872–1884. Available at: https://doi.org/10.1101/gr.216747.116.

Monavarfeshani, A. *et al.* (2023) 'Transcriptomic analysis of the ocular posterior segment completes a cell atlas of the human eye', *Proceedings of the National Academy of Sciences of the United States of America*, 120(34), p. e2306153120. Available at: https://doi.org/10.1073/pnas.2306153120.

Myers, T.A., Chanock, S.J. and Machiela, M.J. (2020) 'LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations', *Frontiers in Genetics*, 11, p. 157. Available at: https://doi.org/10.3389/fgene.2020.00157.

Orozco, L.D. *et al.* (2020) 'Integration of eQTL and a Single-Cell Atlas in the Human Eye Identifies Causal Genes for Age-Related Macular Degeneration', *Cell Reports*, 30(4), pp. 1246-1259.e6. Available at: https://doi.org/10.1016/j.celrep.2019.12.082.

Orozco, L.D. *et al.* (2023) 'A systems biology approach uncovers novel disease mechanisms in age-related macular degeneration', *Cell Genomics*, 3(6), p. 100302. Available at: https://doi.org/10.1016/j.xgen.2023.100302.

Patterson, N., Price, A.L. and Reich, D. (2006) 'Population Structure and Eigenanalysis', *PLOS Genetics*, 2(12), p. e190. Available at: https://doi.org/10.1371/journal.pgen.0020190.

Pedersen, B.S. *et al.* (2020) 'Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches', *Genome Medicine*, 12(1), p. 62. Available at: https://doi.org/10.1186/s13073-020-00761-2.

'Picard toolkit' (2019). Broad Institute. Available at: https://broadinstitute.github.io/picard/.

Pinelli, M. *et al.* (2016) 'An atlas of gene expression and gene co-regulation in the human retina', *Nucleic Acids Research*, 44(12), pp. 5773–5784. Available at: https://doi.org/10.1093/nar/gkw486.

Price, A.L. *et al.* (2006) 'Principal components analysis corrects for stratification in genome-wide association studies', *Nature Genetics*, 38(8), pp. 904–909. Available at: https://doi.org/10.1038/ng1847.

Quinodoz, M. *et al.* (2025) 'De novo and inherited dominant variants in U4 and U6 snRNAs cause retinitis pigmentosa', *medRxiv: The Preprint Server for Health Sciences*, p. 2025.01.06.24317169. Available at: https://doi.org/10.1101/2025.01.06.24317169.

Ratnapriya, R. *et al.* (2019) 'Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration', *Nature Genetics*, 51(4), pp. 606–610. Available at: https://doi.org/10.1038/s41588-019-0351-9.

Rio Frio, T. *et al.* (2008) 'Two trans-acting eQTLs modulate the penetrance of PRPF31 mutations', *Human Molecular Genetics*, 17(20), pp. 3154–3165. Available at: https://doi.org/10.1093/hmg/ddn212.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26(1), pp. 139–140. Available at: https://doi.org/10.1093/bioinformatics/btp616.

Robinson, M.D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome Biology*, 11(3), p. R25. Available at: https://doi.org/10.1186/gb-2010-11-3-r25.

Sayols, S. (2023) ‘rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms’, *microPublication Biology* [Preprint]. Available at: https://doi.org/10.17912/micropub.biology.000811.

Seddon, J.M. *et al.* (2005) ‘The US twin study of age-related macular degeneration: relative roles of genetic and environmental influences’, *Archives of Ophthalmology (Chicago, Ill.: 1960)*, 123(3), pp. 321–327. Available at: https://doi.org/10.1001/archopht.123.3.321.

Sharon, D. *et al.* (2002) ‘Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE)’, *Proceedings of the National Academy of Sciences*, 99(1), pp. 315–320. Available at: https://doi.org/10.1073/pnas.012582799.

Smail, C. *et al.* (2022) ‘Integration of rare expression outlier-associated variants improves polygenic risk prediction’, *The American Journal of Human Genetics*, 109(6), pp. 1055–1064. Available at: https://doi.org/10.1016/j.ajhg.2022.04.015.

Small, K.W. *et al.* (2016) ‘North Carolina Macular Dystrophy Is Caused by Dysregulation of the Retinal Transcription Factor PRDM13’, *Ophthalmology*, 123(1), pp. 9–18. Available at: https://doi.org/10.1016/j.ophtha.2015.10.006.

Snyder, M.P. *et al.* (2020) ‘Perspectives on ENCODE’, *Nature*, 583(7818), pp. 693–698. Available at: https://doi.org/10.1038/s41586-020-2449-8.

Stegle, O. *et al.* (2010) ‘A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression Levels Greatly Increases Power in eQTL Studies’, *PLOS Computational Biology*, 6(5), p. e1000770. Available at: https://doi.org/10.1371/journal.pcbi.1000770.

Strauss, O. (2005) ‘The retinal pigment epithelium in visual function’, *Physiological Reviews*, 85(3), pp. 845–881. Available at: https://doi.org/10.1152/physrev.00021.2004.

Strunz, T. *et al.* (2020) ‘A mega-analysis of expression quantitative trait loci in retinal tissue’, *PLoS genetics*, 16(9), p. e1008934. Available at: https://doi.org/10.1371/journal.pgen.1008934.

Taylor-Weiner, A. *et al.* (2019) ‘Scaling computational genomics to millions of individuals with GPUs’, *Genome Biology*, 20(1), p. 228. Available at: https://doi.org/10.1186/s13059-019-1836-7.

Tenney, A.P. *et al.* (2023) ‘Noncoding variants alter GATA2 expression in rhombomere 4 motor neurons and cause dominant hereditary congenital facial paresis’, *Nature Genetics*, 55(7), pp. 1149–1163. Available at: https://doi.org/10.1038/s41588-023-01424-9.

The 100,000 Genomes Project Pilot Investigators (2021) ‘100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report’, *New England Journal of Medicine*, 385(20), pp. 1868–1880. Available at: https://doi.org/10.1056/NEJMoa2035790.

THE GTEX CONSORTIUM (2020) ‘The GTEx Consortium atlas of genetic regulatory effects across human tissues’, *Science*, 369(6509), pp. 1318–1330. Available at: https://doi.org/10.1126/science.aaz1776.

Turnbull, C. *et al.* (2018) ‘The 100 000 Genomes Project: bringing whole genome sequencing to the NHS’, *BMJ (Clinical research ed.)*, 361, p. k1687. Available at: https://doi.org/10.1136/bmj.k1687.

Vig, A. *et al.* (2020) ‘DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration’, *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 22(12), pp. 2041–2051. Available at: https://doi.org/10.1038/s41436-020-0915-1.

Voigt, A.P. *et al.* (2022) ‘Choroidal endothelial and macrophage gene expression in atrophic and neovascular macular degeneration’, *Human Molecular Genetics*, 31(14), pp. 2406–2423. Available at: https://doi.org/10.1093/hmg/ddac043.

Wakeling, M.N. *et al.* (2022) ‘Non-coding variants disrupting a tissue-specific regulatory element in HK1 cause congenital hyperinsulinism’, *Nature Genetics*, 54(11), pp. 1615–1620. Available at: https://doi.org/10.1038/s41588-022-01204-x.

Wang, K. *et al.* (2025) ‘Molecular mechanisms of CAND2 in regulating SCF ubiquitin ligases’, *Nature Communications*, 16(1), p. 1998. Available at: https://doi.org/10.1038/s41467-025-57065-5.

Wang, S.K. *et al.* (2022) ‘Single-cell multiome of the human retina and deep learning nominate causal variants in complex eye diseases’, *Cell Genomics*, 2(8), p. 100164. Available at: https://doi.org/10.1016/j.xgen.2022.100164.

Wigdor, E.M. *et al.* (2024) ‘Investigating the role of common cis-regulatory variants in modifying penetrance of putatively damaging, inherited variants in severe neurodevelopmental disorders’, *Scientific Reports*, 14(1), p. 8708. Available at: https://doi.org/10.1038/s41598-024-58894-y.

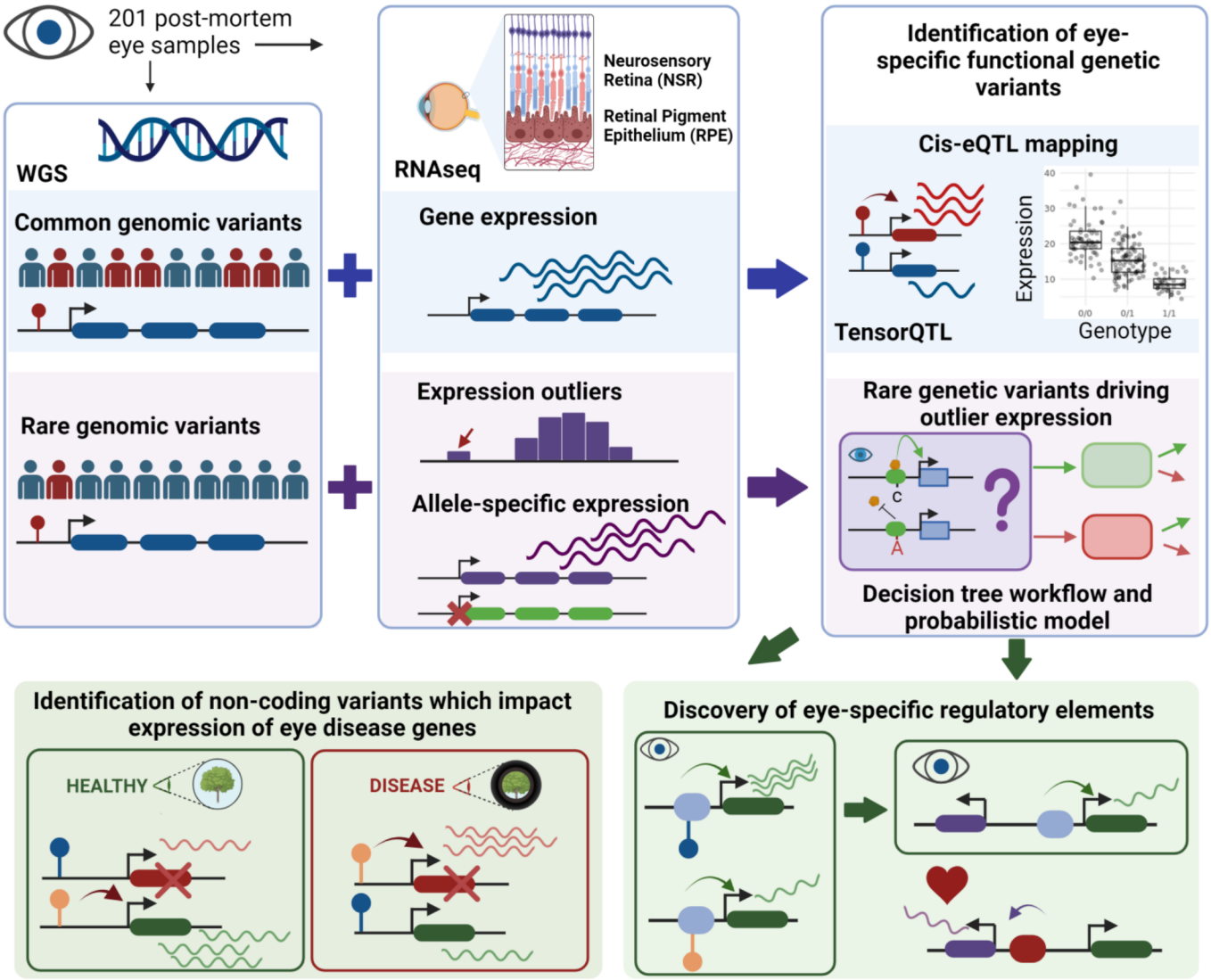
Wong, W.L. *et al.* (2014) ‘Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis’, *The Lancet. Global Health*, 2(2), pp. e106-116. Available at: https://doi.org/10.1016/S2214-109X(13)70145-1.

Wright, A.F. *et al.* (2010) ‘Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait’, *Nature Reviews. Genetics*, 11(4), pp. 273–284. Available at: https://doi.org/10.1038/nrg2717.

Yan, W. *et al.* (2020) ‘Cell Atlas of The Human Fovea and Peripheral Retina’, *Scientific Reports*, 10(1), p. 9802. Available at: https://doi.org/10.1038/s41598-020-66092-9.

Yépez, V.A. *et al.* (2021) ‘Detection of aberrant gene expression events in RNA sequencing data’, *Nature Protocols*, 16(2), pp. 1276–1296. Available at: https://doi.org/10.1038/s41596-020-00462-5.

van Zyl, T. *et al.* (2022) ‘Cell atlas of the human ocular anterior segment: Tissue-specific and shared cell types’, *Proceedings of the National Academy of Sciences of the United States of America*, 119(29), p. e2200914119. Available at: https://doi.org/10.1073/pnas.2200914119.




***Graphical Abstract***

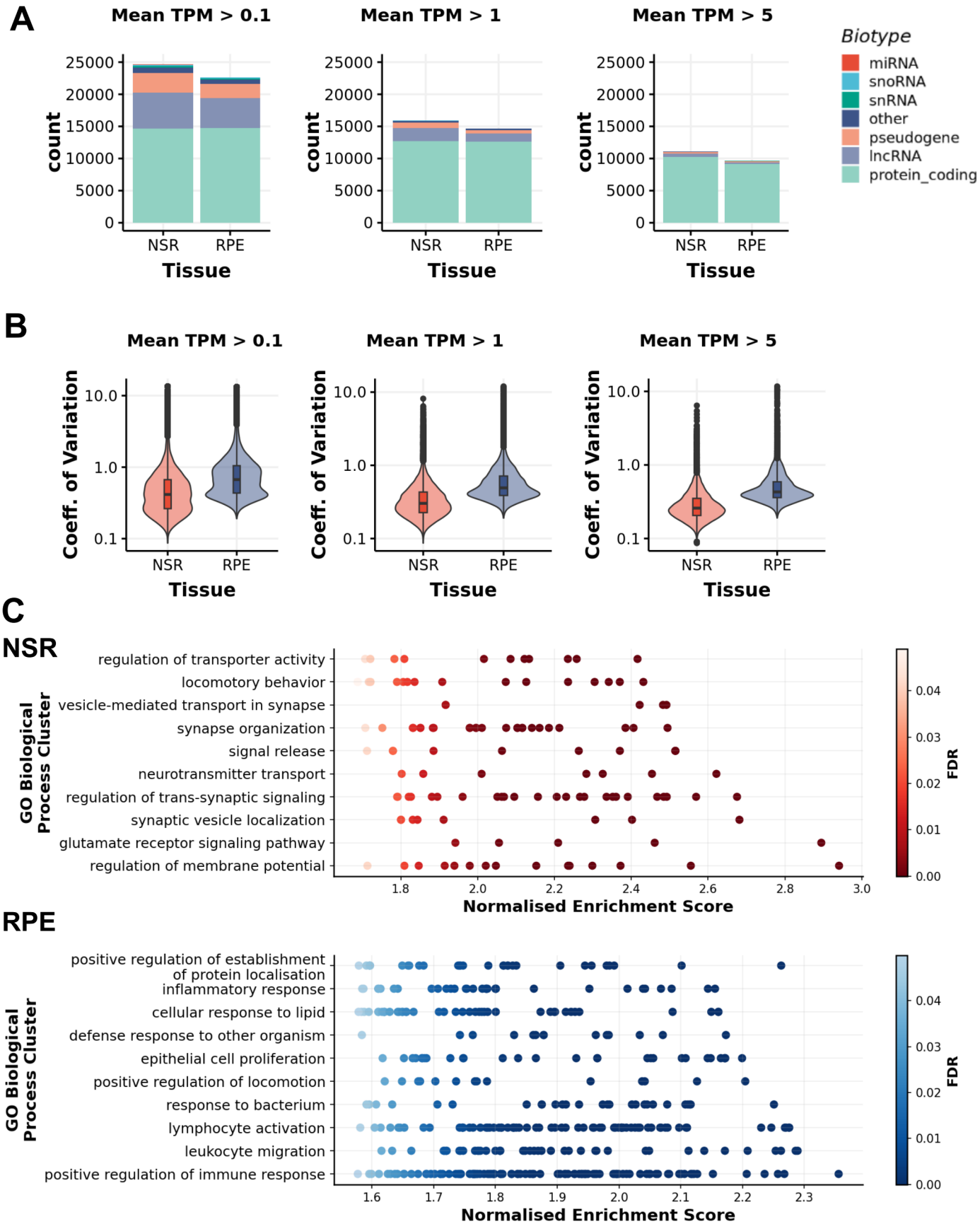
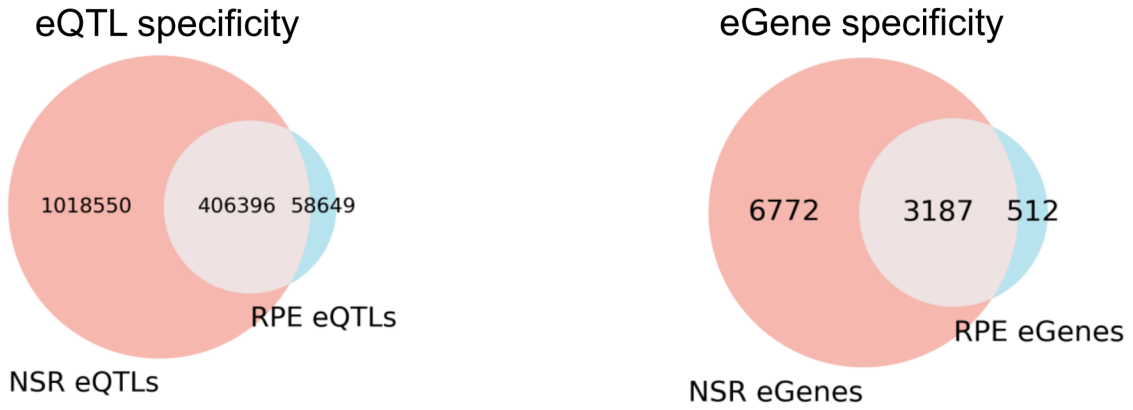
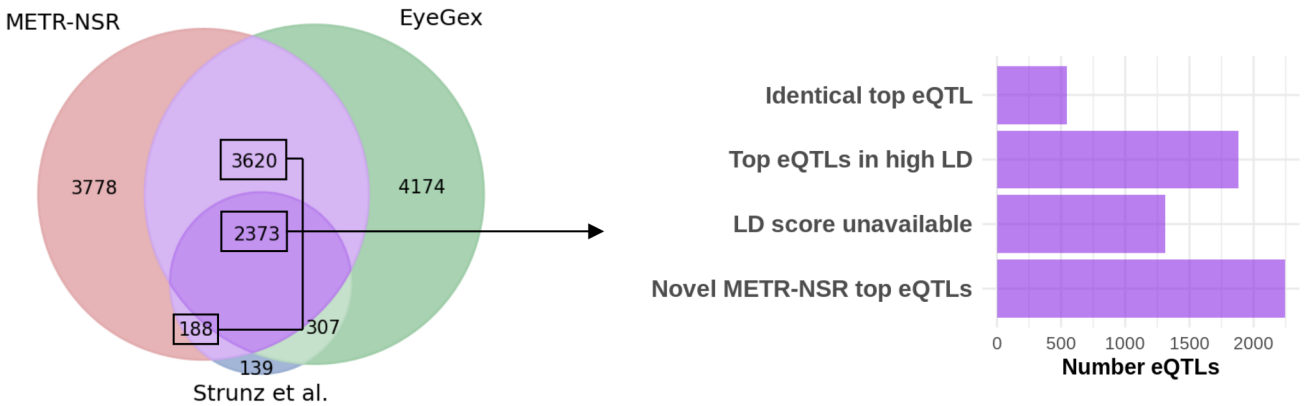



***Figure 1 Tissue-specific transcriptomic data generated for the neurosensory retina (NSR) and the retinal pigment epithelium (RPE)*** *A) Number of genes expressed in the NSR and the RPE at different expression thresholds, classified into different biotypes. B) The coefficient of variation (mean/SD) for all genes expressed at different thresholds in both tissues indicates higher expression variability across samples in the RPE compared to the NSR. C) Top 10 Gene Ontology Biological Process clusters enriched in differentially expressed genes (adj pvalue < 0.01) in the NSR and RPE respectively. Gene Ontology terms (FDR < 0.01) were sorted by enrichment ratio and clustered based on semantic similarity.*

## A. Intersection between METR-NSR eQTLs and METR-RPE eQTLs




## B. Intersection between METR-NSR, EyeGex and Strunz et al. retina eQTLs




## C. Intersection between METR-eQTLs and GTEX eQTLs

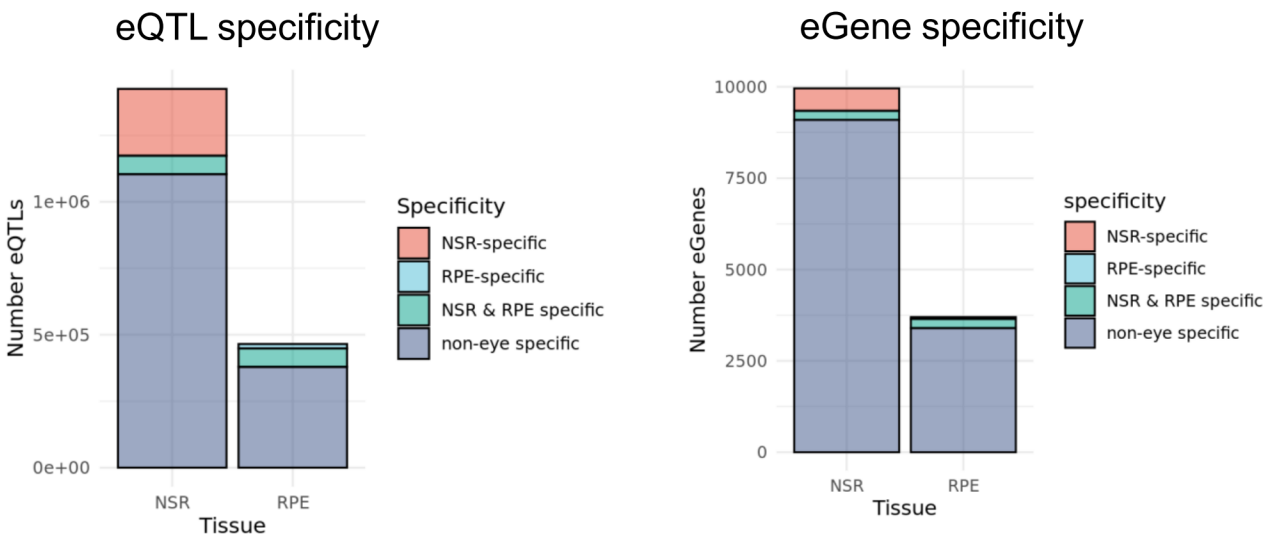



***Figure 2 Intersection between METR-eQTLs and eQTLs from different studies*** *A) Overlap between METR-eQTLs (left) and METR-eGenes (right) identified in the neurosensory retina (NSR) and the retinal pigment epithelium (RPE). B) Intersection between eGenes identified in the neurosensory retina (METR-NSR), EyeGex and Strunz et al. (2020) retina eQTL study. For eGenes which were present in METR-NSR and at least one other study, we compared the top eQTL for each eGene to identify 1) eQTLs which were replicated in an additional study/studies; 2) eGenes where the top eQTL from each study was in high LD with each other (r2 > 0.8) and 3) eGenes where the top NSR hit was novel. For a subset of eQTLs that were tested for LD, the LD score was unavailable. C) Comparison between METR-eQTLs and GTEX eQTLs indicates the number of NSR-specific, RPE-specific, NSR & RPE-specific and non-eye specific eQTLs (left) and eGenes (right) from both tissues in our study*

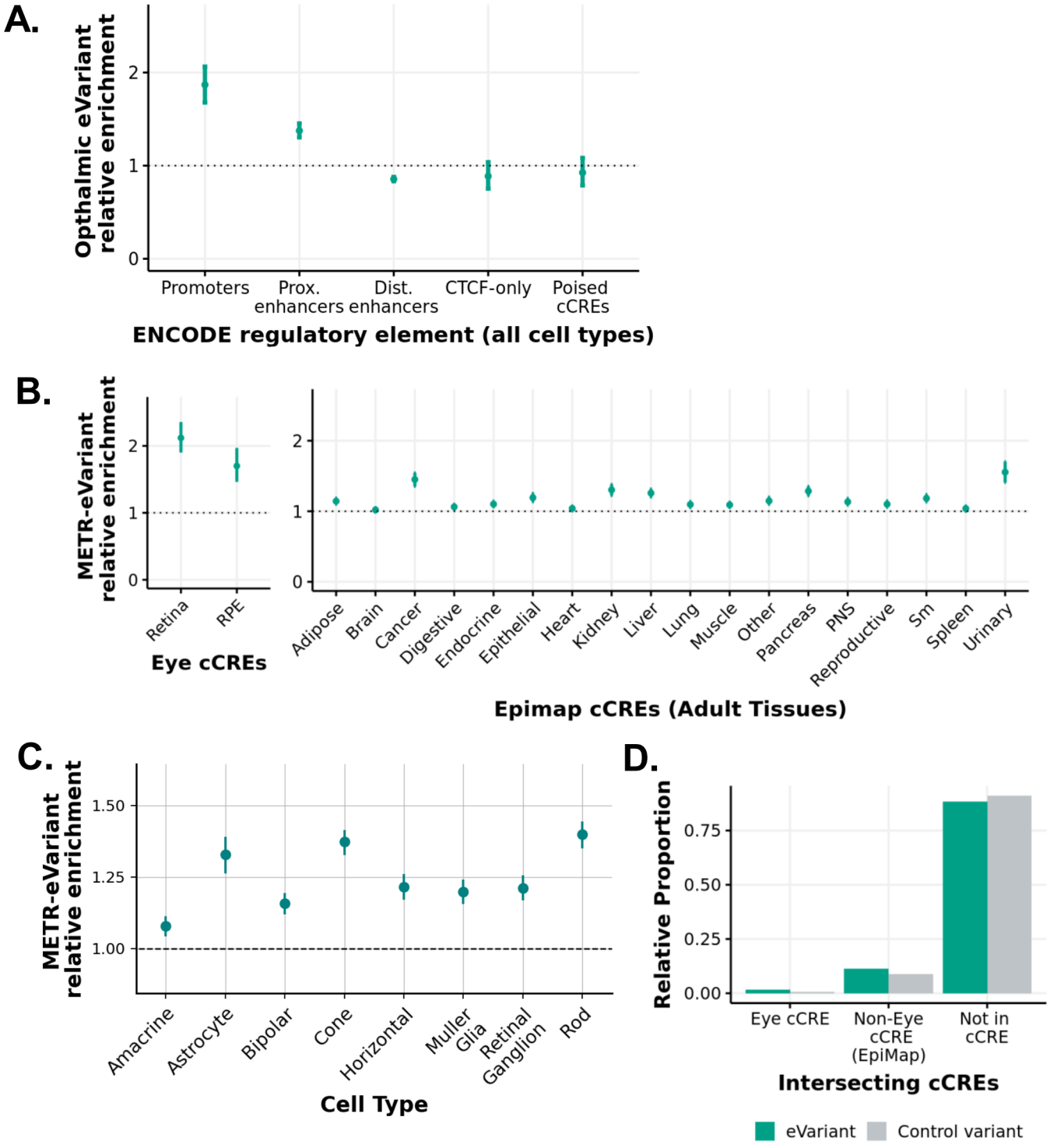



***Figure 3 Enrichment of eye eVariants in previously annotated candidate cis-regulatory elements (cCREs) A)*** *Bootstrapped relative enrichment of eye eVariants which intersect with cCREs from the ENCODE cCRE registry (V3) compared to control variants. Relative enrichment is defined as the ratio of eVariants to control variants which intersect with each element. In each bootstrapping iteration, random subsets of eVariants and control variants (subsets of non-eVariants matched for gnomad AF and gene density) were intersected with each cCRE group. Error bars indicate the 2.5-97.5% confidence intervals.* ***B)*** *Relative enrichment of eye eVariants which intersect with cCREs from retina and RPE (Cherry et al., 2020) and cCREs from adult tissues in EpiMap compared to control variants matched for gnomad AF and gene density.* ***C)*** *Relative enrichment of eVariants which intersect accessible chromatin regions from single cell ATACseq peaks (Wang et al., 2023) from different retina cell types.* ***D)*** *Relative proportion of eVariants and non-eVariants in the NSR and/or RPE which intersect with annotated retina cCREs (Cherry et al., 2020), non-retina cCREs (EpiMap) and neither. Control variants refer to those which were included in eQTL mapping (MAF > 2.5% and AC > 10) but were not associated with any eQTLs in NSR or RPE.*

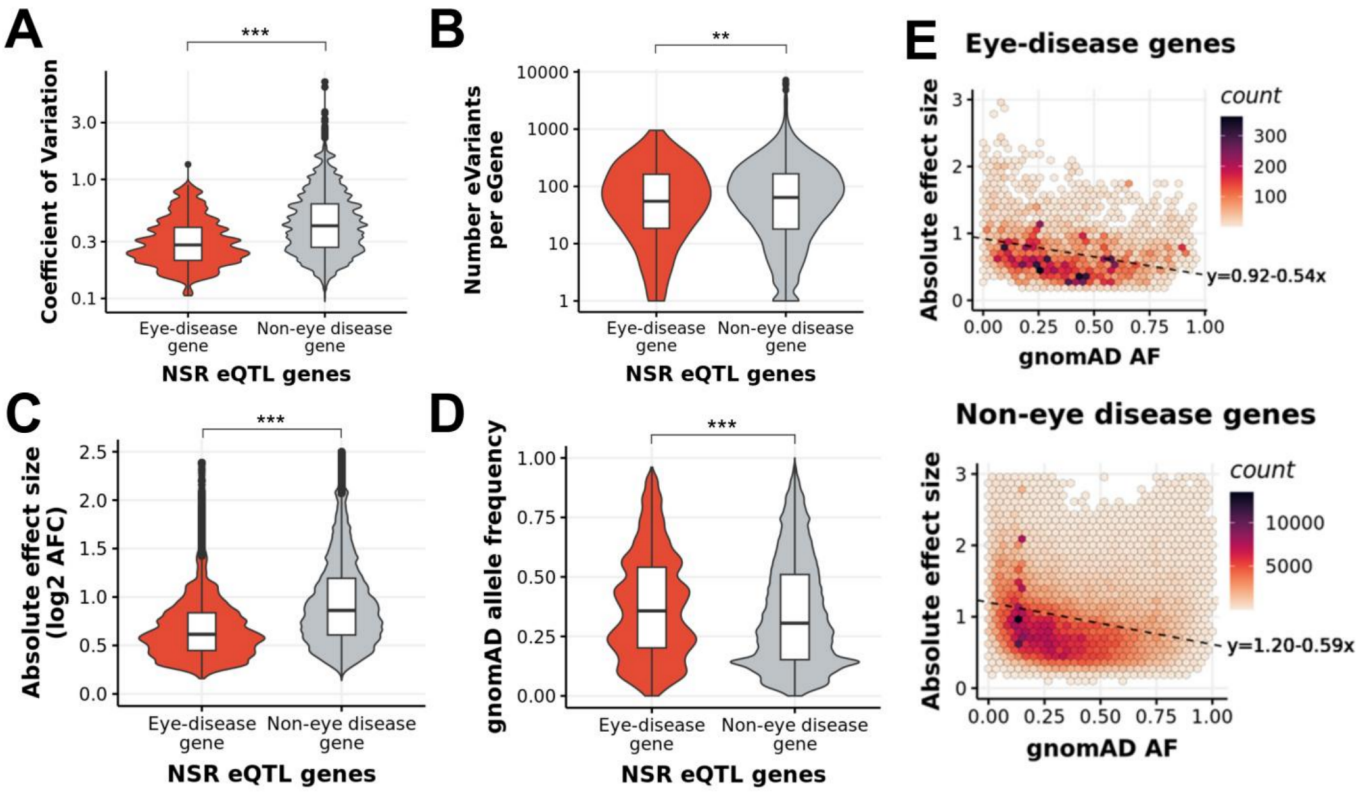



***Figure 4 Differences in NSR eQTLs associated with known rare monogenic eye disease genes and genes which are not attributed to eye diseases (non-eye disease genes) A)*** *Genes associated with eye diseases on EyeG2P have lower coefficients of variation than non-eye disease genes, indicating lower expression variability across samples.* ***B)*** *On average, known eye disease genes have fewer eVariants per gene than non-eye disease genes.* ***C)*** *The impact on gene expression (measured in absolute log2 allelic fold change) of each eQTL associated with a known eye disease gene is lower than the impact of eQTLs associated with non-eye disease genes.* ***D)*** *The allele frequency on gnomAD for the eVariants associated with eye disease genes is higher than non-eye disease genes.* ***E)*** *There is a negative linear relationship between the allele frequency of an eVariant, and the impact on gene expression of the associated eGene (measured in log2(allelic fold change)) for eQTLs associated with eye disease genes (top) and non-eye disease genes (bottom)*

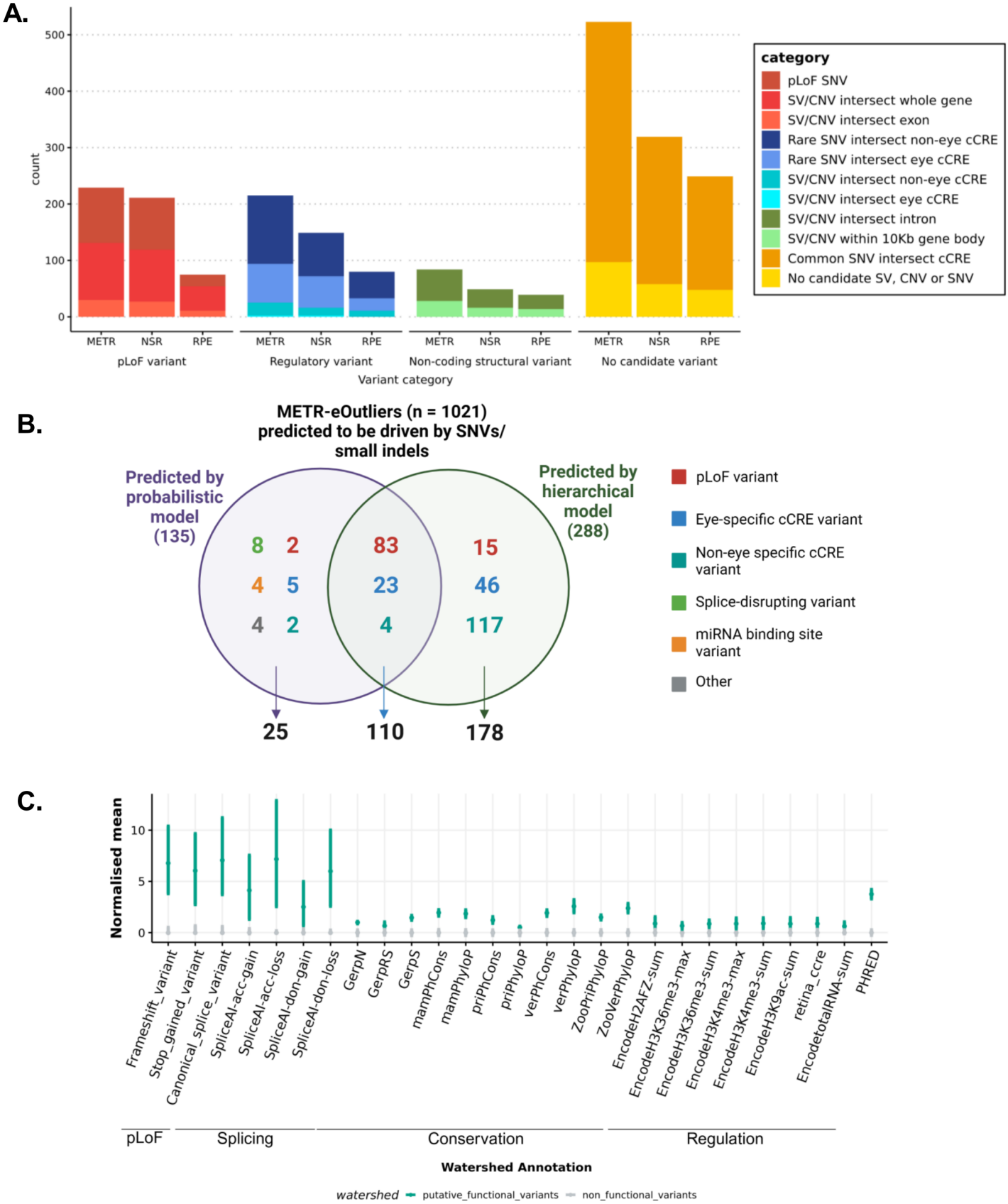



***Figure 5 Identification of rare variants driving METR transcriptomic outliers in neurosensory retina (NSR) and retinal pigment epithelium (RPE). A)*** *A hierarchical workflow to identify candidate variants driving outlier expression identified putative functional variants driving 563 eOutlier events in NSR, RPE or both (METR)* ***B)*** *The Watershed probabilistic model had high concordance with the hierarchical framework for the prioritisation of SNVs and small indels driving eOutliers in NSR and RPE.* ***C)*** *Bootstrapping analysis indicates that variants which were predicted to be driving eOutliers by Watershed were enriched for pLoF variants, splicing variants, variants within regulatory elements (including retina cCREs) and those with high conservation scores compared to those which were not predicted to have a functional impact. Bars indicate the 95% CI.*

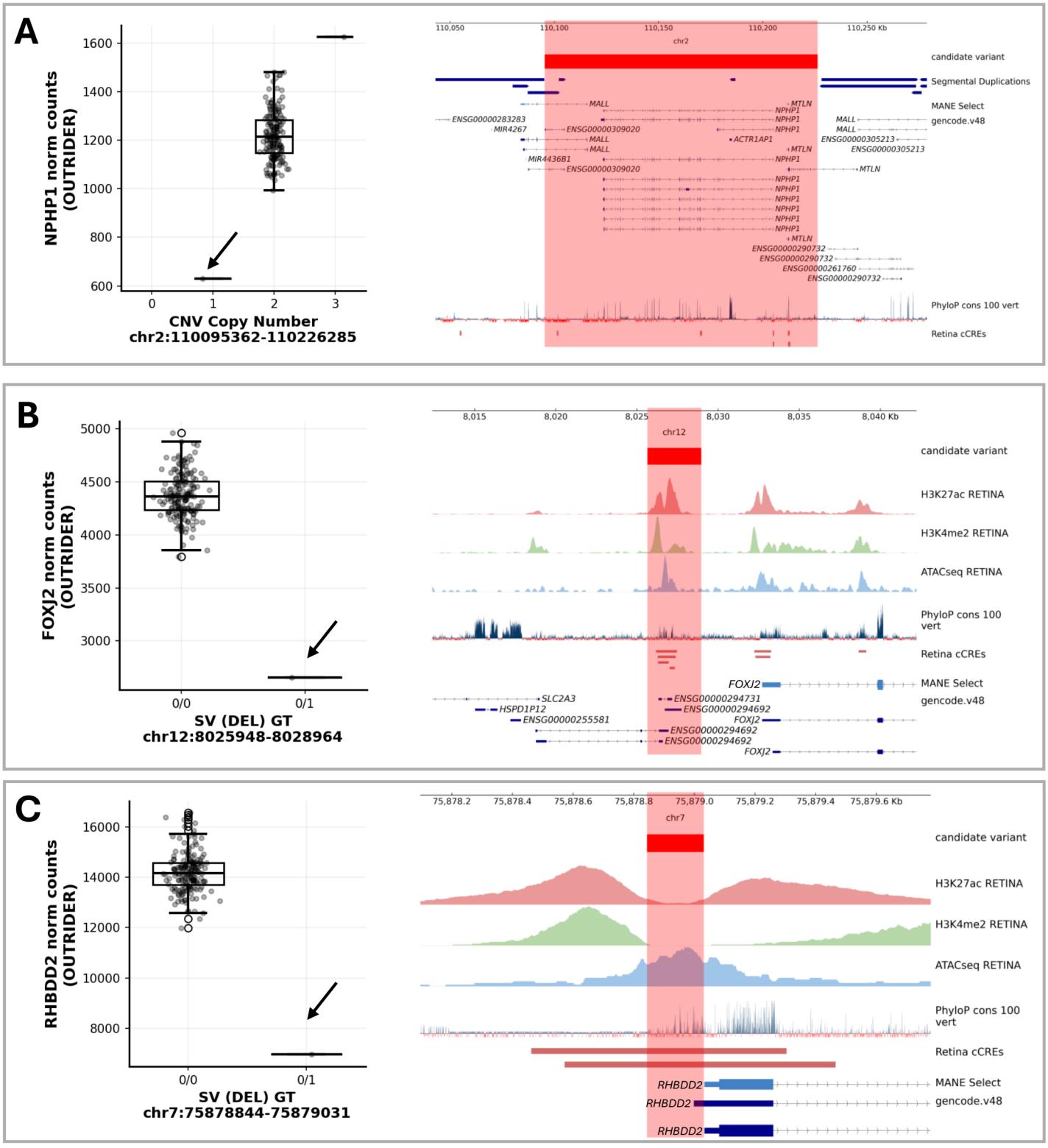



***Figure 6 Candidate structural and copy number variants driving METR transcriptomic outliers in neurosensory retina (NSR).*** *In each caption, the tissue and relative outlier expression profiles calculated through OUTRIDER for individuals are shown, with:* ***(A)*** *copy number states 0,1, 2 and 3; and* ***(B,C)*** *homozygous reference (0/0) and heterozygous alternate (0/1) genotypes. Genome tracks displaying transcript isoforms, evolutionary conservation and candidate cis-regulatory elements (cCREs) identified in Cherry et al are included alongside:* ***(A)*** *segmental duplication regions; and* ***(B,C)*** *epigenomic histone marks and ATACSeq in retina (Cherry et al., 2020). Box and whiskers plots show median values and interquartile ranges, with grey dots indicating normalised count values for single samples, and statistical outliers for each genotype indicated with white dots. In* ***(A)*** CNVs encompassing a known eye disease gene (*NPHP1*) are shown to cause drastic changes in expression in NSR. In **(B)** and **(C)** deletions are shown to impact cCREs proximal to the transcription start sites of genes expressed in NSR: **(B)** 3Kb deletion ~3.5Kb upstream of *FOXJ2;* **(C)** 187bp deletion impacting minimal promoter region of *RHBDD2*.

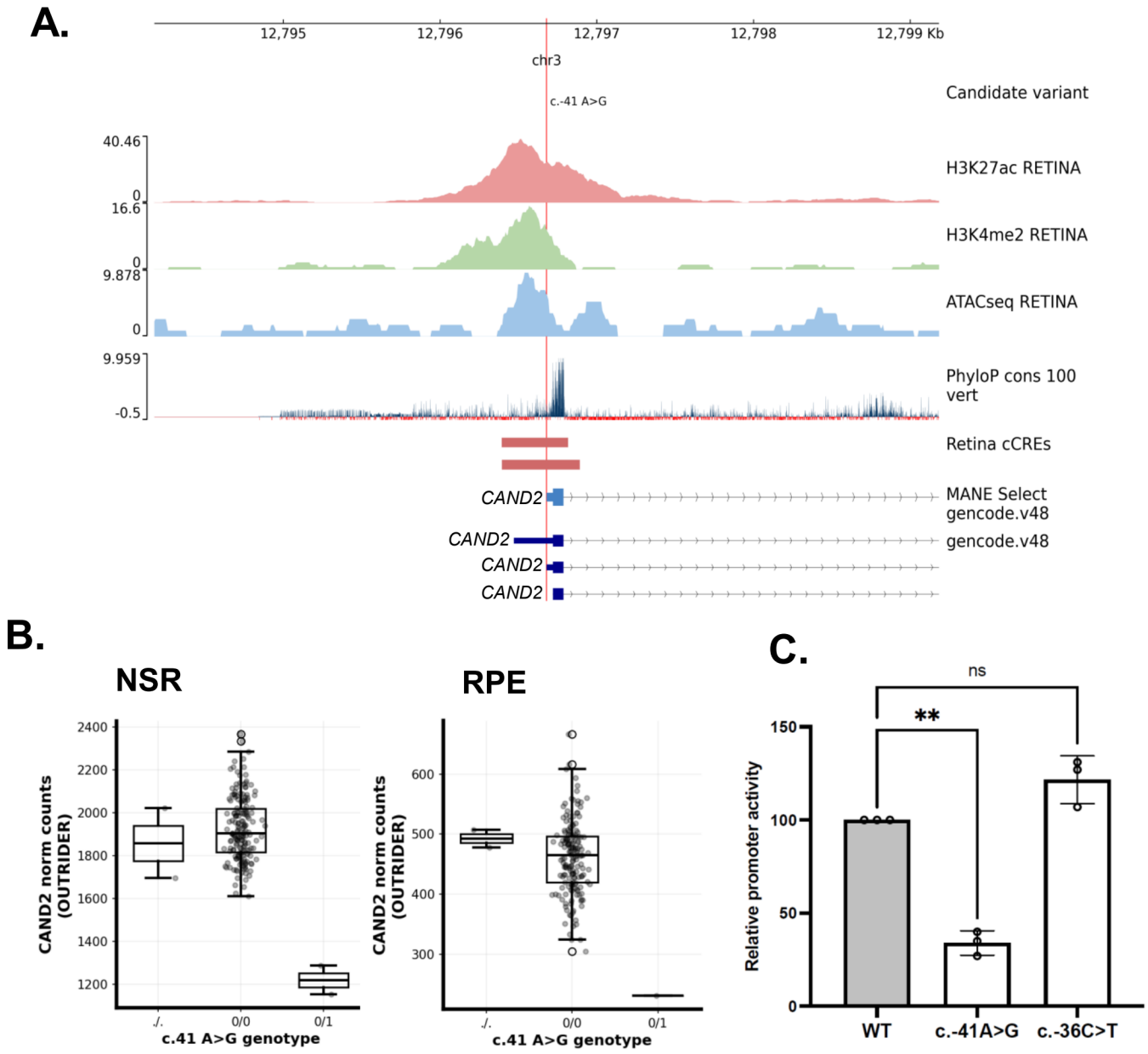



***Figure 7 Candidate rare variant in promoter of CAND2 driving METR transcriptomic outliers in neurosensory retina (NSR) and retinal pigment epithelium (RPE). A)*** *Genome tracks displaying the rare promoter variant identified in CAND2 (NM_001162499.2:c.-41A>G) mapped alongside retina-specific epigenomic peaks and annotated retina-specific cCREs (both generated by Cherry et al. (2020)), evolutionary conservation and GENCODE v48 transcript isoforms* ***B)*** *The outlier expression profiles calculated through OUTRIDER for individuals with missing (./.), homozygous reference (0/0) and heterozygous alternate (0/1) genotypes are shown for NSR and RPE.* ***C)*** *Results from a dual reporter luciferase assay confirm that there is a significant reduction in CAND2 promoter activity in the presence of c.-41A>G (adj p=0.005) compared to the WT promoter and a promoter carrying a common variant in gnomAD (c.-36C>T)*