Single-cell multiome of the human retina and deep learning nominate causal variants in complex eye diseases

Sean K. Wang,[1,2] Surag Nair,[3] Rui Li,[1] Katerina Kraft,[1] Anusri Pampari,[3] Aman Patel,[3] Joyce B. Kang,[4,5] Christy Luong,[1,6] Anshul Kundaje,[3,7] Howard Y. Chang[1,8]

[1] Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA

[2] Department of Ophthalmology, Stanford University School of Medicine, Stanford, CA, USA

[3] Department of Computer Science, Stanford University, Stanford, CA, USA

[4] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

[5] Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA

[6] Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

[7] Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[8] Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

Correspondence to: H.Y.C. at howchang@stanford.edu

ABSTRACT

Genome-wide association studies (GWAS) of eye disorders have identified hundreds of genetic variants associated with ocular disease. However, the vast majority of these variants are noncoding, making it challenging to interpret their function. Here, we present a joint single-cell atlas of gene expression and chromatin accessibility of the adult human retina with >50,000 cells, which we used to analyze noncoding single-nucleotide polymorphisms (SNPs) implicated by GWAS of age-related macular degeneration, glaucoma, diabetic retinopathy, myopia, and type 2 macular telangiectasia. We integrate this atlas with a HiChIP enhancer connectome, expression quantitative trait loci (eQTL) data, and base-resolution deep learning models to predict noncoding SNPs with causal roles in eye disease, assess SNP impact on transcription factor binding, and define their known and novel target genes. Our efforts nominate pathogenic SNP-target gene interactions for multiple vision disorders and provide a potentially powerful resource for interpreting noncoding variation in the eye.

INTRODUCTION

Genome-wide association studies (GWAS) of eye disorders such as glaucoma, myopia, and age-related macular degeneration (AMD) have uncovered hundreds of genetic polymorphisms associated with ocular disease[1–5]. However, the vast majority of variants identified by GWAS reside in noncoding regions of the genome, making it challenging to interpret their function[6]. To better understand how noncoding variants mechanistically contribute to ocular pathology, it would be valuable to map in which cell types their corresponding loci are active. This information would provide novel insights into the cellular biology of genetically complex eye diseases and help nominate specific cell types as targets for therapies.

A recent advance in studying the noncoding genome has been the development of single-cell multiomic technologies such as paired single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq). While scRNA-seq can classify the different cell types of a tissue based on their transcriptional profiles, its combination with scATAC-seq allows for the additional mapping of cell type-specific chromatin accessibility. Together, these techniques can reveal the activity of noncoding DNA elements identified by GWAS and have been used to interrogate risk variants for conditions including Alzheimer's disease, Parkinson's disease, autism spectrum disorder, and autoimmunity[7–9].

Investigations into the noncoding genome have likewise benefitted from analytical innovations, such as the application of convolutional neural network (CNN)-based deep learning to predict the effects of noncoding polymorphisms[10–12]. Progress in this area has recently led to models with resolution down to a single nucleotide, enabling accurate determination of the critical bases

within *cis*-regulatory sequences[8,12,13]. These models offer a validated approach to prioritize noncoding variants with functional relevance and are particularly suitable for tissues in which experimental manipulation is difficult.

Here, we generated a joint scRNA- and scATAC-seq atlas of the adult human retina composed of >50,000 cells and used it to analyze noncoding single-nucleotide polymorphisms (SNPs) implicated by GWAS of five eye diseases: AMD, glaucoma, diabetic retinopathy (DR), myopia, and type 2 macular telangiectasia (MacTel). Layering this atlas with a HiChIP enhancer connectome[14], expression quantitative trait loci (eQTL) data[15], and base-resolution deep learning models[12], we then predicted noncoding SNPs with causal roles in eye disease. Our efforts nominate pathogenic SNP-target gene interactions for multiple vision disorders and provide a potentially powerful resource for interpreting noncoding variation in the eye.

RESULTS

Single-cell multiomics reveal the gene expression and chromatin accessibility landscapes of cell types in the human retina

To generate a single-cell multiome of the human retina, we performed joint scRNA- and scATAC-seq profiling on eight postmortem retinas from four individuals who had no history of eye disease (Supplementary Table 1). Following quality control filtering (Extended Data Fig. 1a-f) and removal of putative doublets (Extended Data Fig. 2a,b), we obtained a total of 51,645 human retinal cells in 22 clusters which we assigned to 13 different cell types (Fig. 1a,b and Extended Data Fig. 2c). These included abundant cell types like rod photoreceptors and Müller glia, as well as rarer cell types such as astrocytes and microglia, which each constituted only

0.4% of profiled cells (Fig. 1c and Supplementary Table 2). Consistent with published scRNA-seq studies of the human retina[16–18], we observed cell type-specific expression of many genes, including *PDE6A* in rod photoreceptors, *GRIK1* in OFF-cone bipolar cells, *RLBP1* in Müller glia, *GRM6* in ON-cone and rod bipolar cells, *PRKCA* in rod bipolar cells, *ARR3* in cone photoreceptors, *GAD1* in GABAergic (GABA-) amacrine cells, *ONECUT1* in horizontal cells, *SLC6A9* in AII- and other glycinergic (gly-) amacrine cells, *NEFL* in retinal ganglion cells, *GJD2* in AII-amacrine cells, *GFAP* in astrocytes, and *C1QA* in microglia (Fig. 1d). In addition, we identified a list of candidate marker genes based on differential expression for each of the 13 cell types (Supplementary Data 1).

Using shared barcodes from joint multiomic profiling, we next assigned scATAC-seq profiles to the 13 cell types characterized above by scRNA-seq. Peak calling performed on scATAC-seq profiles from each cell type combined into pseudo-bulk ATAC replicates uncovered a total of 620,386 chromatin accessibility peaks (Fig. 2a and Supplementary Data 2). These scATAC peaks included >90% of peaks from published bulk ATAC-seq of the human retina (Fig. 2b)[19], indicating that single-cell multiomics can recapitulate bulk ATAC-seq data. Conversely, more than half of scATAC peaks were unique to the single-cell dataset (Fig. 2b), and nearly 40% of scATAC peaks were accessible in only one cell type (Extended Data Fig. 3a). Supporting this, we found 197,826 scATAC marker peaks enriched in a cell type-specific manner (Fig. 2c and Supplementary Data 3), including many located near cell type-specific genes (Fig. 2d).

With these scATAC peaks, we then conducted motif enrichment analysis to predict what transcription factors (TFs) might be active in each cell type (Fig. 3a and Supplementary Data 4).

In accord with published literature, we observed enrichment of binding motifs for TFs with known cell type-specific functions, such as OTX2 in photoreceptors and bipolar cells[20], ONECUT family members in horizontal cells[21], POU4F family members in retinal ganglion cells[22], and SPI1 (PU.1) in microglia[23]. For some TFs, cell type-specific activity was also supported by footprinting analysis of scATAC peaks (Fig. 3b), which revealed motif centers to be protected from Tn5 transposition, consistent with TF occupancy. These data offer a cell type-specific catalog of candidate TFs in the adult retina and may aid our understanding of gene-regulatory networks controlling vision.

Single-cell multiomics uncover the cellular contexts of variants implicated by ocular disease GWAS

Using our single-cell multiome, we sought to better understand risk loci identified by GWAS of complex eye disorders. To this end, we compiled of a list of 1,331 unique index SNPs from the NHGRI-EBI GWAS Catalog representing GWAS hits for five eye diseases: AMD, glaucoma, DR, myopia, and MacTel (Fig. 4a and Supplementary Table 3)[24]. The vast majority (96.5%) of these SNPs localize to noncoding regions of the genome and thus cannot be interpreted with scRNA-seq data alone. We performed linkage disequilibrium (LD) expansion on all index SNPs to include nearby variants with high probability of coinheritance (LD R^2 >0.9 based on phase 3 genotypes from the 1000 Genomes Project) (Extended Fig. 3b)[25]. From this, we obtained a total of 7,034 unique noncoding SNPs in loci associated with eye disorders (Supplementary Data 5).

To determine in which retinal cell types each of the 7,034 SNPs might be active, we overlapped SNP locations with scATAC peaks from our dataset. We found that 1,152 SNPs (16.4%)

overlapped with a scATAC peak (Fig. 4b and Extended Data Fig. 3c), and that most SNP-containing peaks were present in only one or two cell types (Extended Fig. 3d). We next conducted two orthogonal analyses to refine our list of SNPs for those more likely to possess gene regulatory functions (Fig. 4c). First, we identified SNPs in scATAC peaks that were co-accessible with peaks in promoter regions, reasoning that this would select for SNPs in active enhancers. We found 39,552 such promoter peaks in the human retina, 58.3% of which were co-accessible with at least one scATAC peak (Fig. 4d). Leveraging our paired scRNA- and scATAC-seq data, we also searched for SNPs in peaks that had at least one predicted target gene based on same-cell correlations between peak accessibility and gene expression. Using this method, we predicted target genes for 199,055 (32.1%) of the 620,386 scATAC peaks in our dataset (Fig. 4e). For nearly half (44.3%) of these peaks, our predictions differed from the nearest gene on the linear genome (Extended Data Fig. 3e), suggesting that noncoding SNPs do not necessarily regulate their nearest gene.

We identified 241 SNPs in scATAC peaks that were co-accessible with promoter peaks and 374 SNPs that had predicted target genes, with 202 SNPs meeting both criteria. As an example, we examined rs4821699 residing in an intron of *TRIOBP* on chromosome 22. This locus has been implicated in glaucoma by multiple GWAS and encodes a protein thought to regulate cytoskeletal organization[2,26,27]. We observed that rs4821699 was most accessible in retinal ganglion cells (Fig. 4f), the major cell type that undergoes degeneration during glaucoma. Based on correlations with gene expression, the peak containing this SNP was furthermore predicted to target *TRIOBP*. We hypothesize that rs4821699 might therefore play a role in glaucoma by altering *TRIOBP* expression in retinal ganglion cells.

A handful of SNPs associated with eye diseases have been experimentally studied using retinal organoids derived from induced pluripotent stem cells. One such SNP is rs17421627, an index SNP from GWAS of MacTel representing a T-to-G substitution on chromosome 5[5,28]. We determined rs17421627 to be one of only five SNPs for MacTel with a predicted target gene and found the SNP to be most accessible in Müller glia and astrocytes (Fig. 4f). Using linked gene expression data, we also predicted rs17421627 to act on *LINC00461*, a long noncoding RNA. Consistent with these predictions, deletion of the locus containing rs17421627 in human retinal organoids has been shown to significantly downregulate *LINC00461* with the strongest effect in Müller glia[29]. These examples illustrate how single-cell multiomics can reveal the cellular targets of noncoding variants in the retina and nominate how they might contribute to eye disorders.

Integration of single-cell multiome with HiChIP and eQTL data validates SNP-target gene predictions

To further prioritize our list of SNPs, we combined our data with two complementary methods for identifying functional SNP-gene interactions genome-wide (Fig. 5a). We first performed HiChIP for acetylated histone H3 lysine 27 (H3K27ac), a mark of active enhancers and promoters[30], to characterize the three-dimensional (3D) enhancer "connectome" of the human retina (Extended Data Fig. 4a,b)[14,31]. We uncovered 16,692 loop anchors connected by 9,670 HiChIP loops, including several linking regions of chromatin accessibility to the transcription start sites (TSSs) of cell type-specific genes (Extended Data Fig. 4c). Of these loops, >95% overlapped with a scATAC peak in both anchors, and >99% overlapped with a peak in at least one anchor (Fig. 5b). This result shows that accessible chromatin sites identified in scATAC-seq

data possess biochemical characteristics of active enhancers and supports their connection to target genes. We additionally analyzed our list of SNPs using published human retina eQTL data from the Eye Genotype Expression (EyeGEx) database[15]. For >90% of SNPs in scATAC peaks, retina eQTL data was available (Fig. 5c), enabling genes whose mRNA expression in the human retina changed with specific SNPs to be identified at the bulk tissue level.

We found 187 disease-associated SNPs in scATAC peaks that were linked to a gene by a H3K27ac HiChIP loop. These included rs9966620, the top SNP from a GWAS of DR representing a G-to-A transition in an intron of *TTC39C* on chromosome 18[32]. Using our multiome, we determined that the scATAC peak containing rs9966620 was most accessible in rods (Fig. 5d). However, this peak also correlated with the expression of multiple target genes, hampering efforts to interpret how the SNP might function. Incorporating our HiChIP data, we were able to locate a 3D loop connecting rs9966620 with a region 75 kilobases (kb) upstream. This region intersected the TSS of only one gene, *TTC39C-AS1*, suggesting that rs9966620 may modulate DR risk by interacting with *TTC39C-AS1* in rods.

We additionally detected 596 disease-associated SNPs in scATAC peaks that were significantly associated with a gene by eQTL analysis. One example was rs2730260, a SNP in an intron of *VIPR2* that has been implicated in myopia[33]. This locus encodes one of two known receptors for vasoactive intestinal peptide (VIP), a signaling molecule involved in visual processing[34]. We found that rs2730260 resided in a chromatin accessibility peak specific to Müller glia that again had multiple predicted target genes (Fig. 5d). This ambiguity was clarified by retina eQTL data, which showed that variation at rs2730260 significantly correlated with the expression of only

VIPR2 (Fig. 5e), supporting this gene as the SNP's primary target. Integration of eQTL data similarly improved our interpretation of rs66475830 on chromosome 6 in the *FRK-NT5DC1-COL10A1* risk locus for AMD[35,36]. This region contains nearly 20 genes within a span of a megabase (Mb), making it particularly difficult to functionally annotate GWAS hits. From our single-cell data, we determined rs66475830 to be accessible in amacrine and horizontal cells and predicted *TSPYL1* and *TSPYL4* as target genes (Fig. 5d). Retina eQTL analysis revealed that variation at this position was significantly associated with *TSPYL4* expression, but not that of other nearby genes (Fig. 5e), nominating *TSPYL4* as the effector gene of rs66475830.

Lastly, we identified many SNP-target gene relationships supported by both HiChIP and eQTL data, such as rs77272443 and rs4102217 located in risk loci for myopia and glaucoma, respectively[37,38]. For both of these SNPs, HiChIP and eQTL analyses again refined target gene predictions (Extended Data Fig. 5a,b), demonstrating how the combination of single-cell multiomics with other assays can enhance interpretation of noncoding variants in eye disease.

Integration of single-cell multiome with base-resolution deep learning nominates functional mechanisms for disease-associated SNPs

CNN-based deep learning models have proven capable of discerning disease-associated SNPs from other noncoding variants[8,10,11]. As a final method to prioritize SNPs in our dataset, we therefore trained CNNs derived from the BPNet architecture on scATAC-seq profiles for each of the 13 retinal cell types (Fig. 6a and Extended Data Fig. 6a,b)[12]. At each SNP region, we compared the projected per-base change in chromatin accessibility between reference and alternate alleles using models specific to the different cell types. These calculations allowed us to

identify "high effect" SNPs, which we defined as SNPs predicted to cause a statistically significant (false discovery rate <0.01) absolute log2 fold change of allele-specific read counts >0.5 in local chromatin accessibility in any cell type.

We found 23 SNPs (2.0%) residing in scATAC peaks that qualified as high effect, a greater percentage than among index SNPs, LD expanded SNPs, random SNPs matched for GC content, and random SNPs residing in scATAC peaks (Fig. 6b and Supplementary Data 7). One of the top scoring SNPs was rs1532278, an index SNP associated with myopia and residing in an intron of *CLU* on chromosome 8[3]. Our atlas predicted rs1532278 to regulate *CLU*, a notion reinforced by eQTL data, and determined the SNP to be accessible in nine of 13 retinal cell types (Fig. 6c,d). Despite this, base-resolution models projected a T-to-C transition at rs1532278 to alter chromatin accessibility only in Müller glia, specifically by disrupting the motif of a homeodomain TF. Our findings suggest that even though rs1532278 is accessible across multiple cell types, its functional impact in the retina might be restricted to Müller glia due to a cell type-specific homeodomain TF. We speculate that this TF could be LHX2 given its robust expression in Müller glia by both our scRNA-seq data (Fig. 6e) as well as data from animal models[39].

Another high effect SNP was rs1874459 located in an intron of *CDH11* on chromosome 16, a locus implicated by multiple GWAS for glaucoma[2,26]. Using our multiome, we found rs1874459 to be most accessible in rod bipolar cells and predicted *CDH11* as one of its target genes, an idea supported by eQTL data (Fig. 6c,f). Incorporating base-resolution models, we then determined that the G-to-C transversion represented by rs1874459 introduced a new basic helix-loop-helix (bHLH) domain, which was expected to increase accessibility in rod bipolar, OFF-cone bipolar,

ON-cone bipolar, gly-amacrine, and AII-amacrine cells. Of the bHLH TFs, members of the neuroD and neurogenin families in particular were predicted by motif analysis to be significantly enriched in these five cell types (Fig. 3a and Supplementary Data 4). We thus compared all neuroD and neurogenin family members using our scRNA-seq data, which revealed only NEUROD4 to be specific to bipolar and amacrine cells (Fig. 6g), consistent with its role in specifying these cell types during development[40,41]. Together, our results suggest that rs1874459 may act on *CDH11* in bipolar and amacrine cells by creating a new bHLH domain recognized by NEUROD4.

DISCUSSION

In this study, we applied single-cell multiomics, HiChIP, eQTL analysis, and base-resolution deep learning to the human retina to decipher the role of noncoding risk variants in five eye diseases. Integrating these methods allowed us to predict gene and cellular targets in the retina for hundreds of SNPs and nominate dozens as pathogenic and meriting functional validation. From an initial list of >7,000 noncoding SNPs, we identified 1,152 located in chromatin accessibility peaks. We subsequently focused on SNPs 1) that were co-accessible with a promoter, 2) whose accessibility correlated with the expression of a nearby gene, 3) that were linked to a gene in 3D space by a H3K27ac HiChIP loop, 4) that demonstrated significant association with a gene based on retina eQTL data, and 5) that were predicted to alter local chromatin accessibility as determined by base-resolution models. We propose that SNPs meeting most or all of these criteria (Extended Data Fig. 7a-e and Supplementary Data 5) be prioritized in future validation efforts.

Our findings build upon recent works that used primarily fetal tissue and stem cell-derived organoids to map cell type-specific chromatin accessibility in the human retina[29,42]. Datasets from these studies offer a rich resource for decoding retinal development, but might not fully recapitulate the biology of the mature retina, making them potentially less suitable for studying eye disorders that present later in life. Here, we not only generated a single-cell multiome of the adult human retina to pinpoint cellular targets for disease-associated SNPs, but also combined it with multiple orthogonal analyses to define putative SNP-target gene interactions. By performing base-resolution deep learning, we were further able to uncover insights not readily apparent from single-cell, HiChIP, and eQTL data, such as the predicted impact of SNPs on TF binding and the directionality of these effects. To facilitate its use, our atlas is publicly available at https://eyemultiome.su.domains/.

Finally, it should be noted that the majority of SNPs we examined did not overlap with any chromatin accessibility peaks, suggesting that they were not active in the retina. We hypothesize that many of these unassigned SNPs may instead function in other parts of the eye and thus could not be captured by our analysis. For instance, although the neural retina is damaged in AMD and DR, the retinal pigment epithelium and vasculature, respectively, are thought to be the primary sites of pathology[43,44]. In glaucoma, the trabecular meshwork and ciliary body can modulate disease severity as evidenced by treatments that act on these tissues[45]. Likewise, the choroid and sclera may be involved in myopia given that they elongate alongside the retina with increasing nearsightedness[46]. Multiomic characterization of these additional ocular regions would enable a more complete understanding of how noncoding SNPs contribute to vision disorders.

METHODS

Human tissues

Postmortem adult human retinas were procured from consented donors by Lions VisionGift (Portland, OR, USA) or Lions Gift of Sight (St Paul, MN, USA) under protocols approved by the Eye Bank Association of America. None of the donors had a history of ocular disease. De-identified retinas were flash-frozen in liquid nitrogen with a maximum death-to-preservation interval of 12 hours and shipped to Stanford University for processing.

Nuclei isolation

Nuclei were isolated from frozen retinas using the Omni-ATAC protocol (https://doi.org/10.17504/protocols.io.6t8herw)[47]. Briefly, tissues were Dounce homogenized in cold homogenization buffer containing 0.3% IGEPAL CA-630 in the presence of protease and RNase inhibitors to release nuclei from frozen cells. Nuclei were subsequently purified via iodixanol gradient centrifugation and washed with ATAC resuspension buffer containing RNase inhibitor and 0.1% Tween-20 before permeabilization following the 10x Genomics demonstrated protocol for complex tissues (CG000375, Rev. B). After resuspension in diluted nuclei buffer, nuclei were counted using a manual hemocytometer to achieve a targeted nuclei recovery of 10,000 nuclei per sample.

scRNA- and scATAC-seq library generation

Joint scRNA- and scATAC-seq libraries were prepared using the 10x Genomics Single Cell Multiome ATAC + Gene Expression kit according to manufacturer's instructions. Libraries were

sequenced with paired-end 150-bp reads on an Illumina NovaSeq 6000 to a target depth of 250 million read pairs per sample.

scRNA- and scATAC-seq data preprocessing and quality control

Demultiplexed scRNA- and scATAC-seq fastq files were inputted into the Cell Ranger ARC pipeline (version 2.0.0) from 10x Genomics to generate barcoded count matrices of gene expression and ATAC data. For each sample, count matrices were loaded in ArchR and selected for barcodes that appeared in both the scRNA-seq and scATAC-seq datasets[48]. Samples in ArchR were quality control filtered for nuclei with 200-50,000 RNA transcripts, <1% mitochondrial reads, <5% ribosomal reads, TSS enrichment >6, and >2,500 ATAC fragments. Quality control filtered nuclei subsequently underwent automated removal of doublets using the filterDoublets function in ArchR, which identifies and removes the nearest neighbors of simulated doublets[48].

scRNA-seq data analysis

scRNA-seq data from nuclei remaining after quality control filtering and automated removal of doublets were analyzed using Seurat (version 3.1.5)[49]. After merging all preprocessed samples into a single Seurat object, gene expression counts were normalized using the NormalizeData function, scaled using the ScaleData function, and batch corrected using Harmony[50]. Graph-based clustering was then performed on the Harmony-corrected data using the top 20 principal components at a resolution of 0.5. Cluster identities were manually annotated based on the expression of genes from published scRNA-seq studies of the human retina[16–18]. Marker genes for each cluster were additionally identified using the FindAllMarkers function with a minimum

fraction of 0.5 and a log2 fold change of 1 (Supplementary Data 1). Clusters expressing

canonical marker genes from different cell types were designated as putative doublets and

excluded, after which re-clustering was performed using the same parameters. Clusters with no

detected marker genes were also excluded, after which the dataset was also re-clustered. Clusters

in the final dataset representing subpopulations of the same cell type were grouped together for

downstream analyses.

scATAC-seq data analysis

scATAC-seq data were analyzed using ArchR (version 1.0.1) based on barcoded cell type

identities from scRNA-seq[48]. For each cell type, pseudo-bulk ATAC replicates were created

using the addGroupCoverages function with default parameters, which generated between two to

five replicates depending on how many cells of that type were present in each sample. Chromatin

accessibility peaks on chromosomes 1-22 and X and outside of blacklist regions were then called

using the addReproduciblePeakSet function and MACS2[51,52], with scATAC peaks for each cell

type defined as those present in at least two pseudo-bulk ATAC replicates (Supplementary Data

2). Marker peaks were identified using the getMarkerFeatures function with a log2 fold change

≥1 and false discovery rate ≤0.01 as determined by Wilcoxon pairwise comparisons

(Supplementary Data 3). Promoter peaks were defined as scATAC peaks within 2,000 bp

upstream or 100 bp downstream of a TSS, and peaks co-accessible to promoter peaks were

identified using the getCoAccessibility function with a correlation cutoff of 0.3 and resolution of

1. Predicted target genes for each scATAC peak were generated using the getPeak2GeneLinks

function integrating barcode-matched RNA expression data from scRNA-seq with a correlation

cutoff of 0.3 and resolution of 1. Nearest genes were determined using the BEDTools closest function based on gene annotations from TxDb.Hsapiens.UCSC.hg38.knownGene[53].

Bulk ATAC-seq data analysis

Bulk ATAC-seq analysis was performed on published ATAC-seq data from five healthy human retinas[19]. After adapter trimming, fastq files were mapped to the hg38 genome using Bowtie2 and filtered to remove PCR duplicates and retain reads from only chromosomes 1-22 and X[54]. Peak calling was then conducted individually on each sample using MACS2[51], followed by exclusion of peaks in blacklist regions[52]. Peak calls present in at least two of the five retinas were included in the bulk ATAC-seq peak set.

Sequencing tracks

Sequencing tracks of chromatin accessibility were generated in ArchR using the plotBrowserTrack function and were normalized by the total number of reads in TSS regions[48]. All data were aligned and annotated to the hg38 reference genome.

Motif enrichment analysis

TF motif enrichment analysis was performed on scATAC peaks using the peakAnnoEnrichment function in ArchR with default parameters based on position frequency matrices from JASPAR 2018 (Supplementary Data 4)[48,55]. Footprinting analysis of TFs was conducted using the getFootprints function in ArchR[48]. To correct for Tn5 insertion bias, footprinting signals were divided by the Tn5 insertion signal prior to plotting.

SNP selection and LD expansion

Index SNPs implicated in AMD, glaucoma, DR, myopia, or MacTel and located on chromosomes 1-22 and X were collected from the NHGRI-EBI GWAS Catalog, a curated collection of human GWAS[24]. LD expansion was then performed using LDlinkR to add any SNPs in LD with each index SNP[56], defined as a LD R^2 value >0.9 in the phase 3 genotypes of the 1000 Genomes Project[25]. LD expanded SNPs were filtered to exclude variants in coding regions based on annotations in dbSNP to obtain the final set of noncoding SNPs (Supplementary Data 5)[57]. A list of all GWAS used in this study is provided in Supplementary Table 3.

HiChIP library generation

H3K27ac HiChIP libraries were prepared as previously reported with minor modifications[14]. Briefly, following isolation of nuclei from frozen retinas as described above, ~8 million nuclei from each sample were washed with nuclei isolation buffer from the diploid chromatin conformation capture (Dip-C) protocol and fixed with 2% paraformaldehyde at room temperature for 10 minutes[58]. Fixed nuclei were then washed twice with cold 1% bovine serum albumin in phosphate-buffered saline before resuspension in 0.5% sodium dodecyl sulfate and resumption of the published HiChIP protocol. Digestion was performed using the MboI restriction enzyme, and sonication was conducted using a Covaris E220 with 5 duty cycles, peak incident power of 140, and 200 cycles per burst for 4 minutes. The ab4729 ChIP validated antibody from Abcam was used to target H3K27ac. HiChIP libraries were sequenced with paired-end 75-bp reads on either an Illumina HiSeq 400 or Illumina NextSeq 550.

HiChIP data analysis

HiChIP sequencing files were initially processed using the HiC-Pro pipeline (version 2.11.0) to remove duplicate reads, assign reads to MboI restriction fragments, filter for valid interactions, and generate binned interaction matrices[59]. Filtered read pairs from HiC-Pro were subsequently converted into .hic files and inputted into HiCCUPS from the Juicer pipeline to call loops (Supplementary Data 6)[60]. HiChIP interaction maps depicting all valid interactions identified by HiC-Pro were visualized using Juicebox[61].

eQTL analysis

Retina eQTL data were obtained from the Eye Genotype Expression (EyeGEx) database[15]. Each of the 1,152 SNPs overlapping with a scATAC peak was searched in the database and the nominal P value of any gene associations with that SNP noted. Adjusted P values were calculated by multiplying the nominal P value by the number of SNP-gene pairs tested for that SNP. Interactions with an adjusted P value <0.05 were considered significant.

Deep learning model training

scATAC-seq reads from the Cell Ranger ARC pipeline were aggregated by cell type to generate cell type-specific fragments files. The fragments files were converted to BigWig tracks of base-resolution Tn5 insertion sites with an +4/-4 shift to account for Tn5 shift. For each cell type, in addition to the peak regions, we selected an equal number of non-peak regions that were matched for GC content in their peaks. We then trained cell type-specific BPNet models to predict the log counts and base-resolution Tn5 insertion profiles as previously reported[8,12]. Briefly, the BPNet model takes as input a 2,114 bp one-hot encoded input sequence and predicts the ATAC-seq

profile and log counts in a 1,000 bp window centered at the input sequence. Following BPNet formulation, we used a multinomial negative log likelihood (MNLL) for the profile output of the model and a mean square error (MSE) loss for the log counts output of the model. The relative loss weight used for the counts loss was 0.1 times the mean total counts per region. During each epoch, training examples were jittered by up to 500 bp on either side and a random half of the sequences were reverse complemented. Each batch contained a 10:1 ratio of peaks to non-peak regions. Five models were trained for each cell type corresponding to five disjoint training folds. Model training was performed using Keras/Tensorflow 2. Code used for model training is available at https://github.com/kundajelab/retina-models. Models are available at https://doi.org/10.5281/zenodo.6330053.

SNP scoring with BPNet

To score LD expanded SNPs associated with eye disease, we centered the input window at the SNP and obtained the log2 fold change in predicted counts between the reference and alternate alleles for each cell type-specific model. We averaged the log2 fold change over the five model folds for each SNP and cell type. To obtain P values, we performed one-sided Poisson tests of the predicted alternate allele count with the rate parameter set to the predicted reference allele count (counts averaged over five folds). For each SNP, we combined P values across cell types with Fisher's method and performed Benjamini-Hochberg correction. SNPs with an absolute fold-averaged log2 fold change >0.5 and false discovery rate <0.01 were assigned putative "high effect" annotation. To obtain a background set, random noncoding SNPs were chosen by shuffling a list of all SNPs from the 1000 Genomes Project[25], filtering out coding regions, and selecting the first 10,000 entries. Only random SNPs localized to chromosomes 1-22 and X were

then retained, leaving 9,984 background SNPs. Background SNPs had similar GC content as disease-associated LD expanded SNPs (51% versus 52%) and were scored as described above. Base importance tracks were visualized using Logomaker[62].

Data availability

Raw and processed scRNA-seq, scATAC-seq, and HiChIP data from this study have been uploaded to GEO under the accession number GSE196235. A web page summarizing these data is additionally available at https://eyemultiome.su.domains/. BPNet models are available at https://doi.org/10.5281/zenodo.6330053. All data generated in this study are available upon reasonable request.

Code availability

Code used for BPNet model training is available at https://github.com/kundajelab/retina-models. All other custom code used in this study is available upon request.

REFERENCES

1. Fritsche, L. G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–43 (2016).

2. MacGregor, S. *et al.* Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. *Nat. Genet.* **50**, 1067–1071 (2018).

3. Tedja, M. S. *et al.* Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat. Genet.* **50**, 834–848 (2018).

4. Grassi, M. A. *et al.* Genome-wide meta-analysis for severe diabetic retinopathy. *Hum. Mol. Genet.* **20**, 2472–81 (2011).

5. Scerri, T. S. *et al.* Genome-wide analyses identify common variants associated with macular telangiectasia type 2. *Nat. Genet.* **49**, 559–567 (2017).

6. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* **11**, 424 (2020).

7. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).

8. Trevino, A. E. *et al.* Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053-5069.e23 (2021).

9. King, H. W. *et al.* Integrated single-cell transcriptomics and epigenomics reveals strong germinal center-associated etiology of autoimmune risk loci. *Sci. Immunol.* **6**, eabh3768 (2021).

10. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–4 (2015).

11. Kelley, D. R., Snoek, J. & Rinn, J. L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **26**, 990–9 (2016).

12. Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021).

13. Kim, D. S. *et al.* The dynamic, combinatorial cis-regulatory lexicon of epidermal differentiation. *Nat. Genet.* **53**, 1564–1576 (2021).

14. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).

15. Ratnapriya, R. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).

16. Menon, M. *et al.* Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat. Commun.* **10**, 4902 (2019).

17. Lukowski, S. W. *et al.* A single-cell transcriptome atlas of the adult human retina. *EMBO J.* **38**, e100811 (2019).

18. Yan, W. *et al.* Cell Atlas of The Human Fovea and Peripheral Retina. *Sci. Rep.* **10**, 9802 (2020).

19. Wang, J. *et al.* ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat. Commun.* **9**, 1364 (2018).

20. Wang, S., Sengel, C., Emerson, M. M. & Cepko, C. L. A gene regulatory network controls the binary fate decision of rod and bipolar cells in the vertebrate retina. *Dev. Cell* **30**, 513–27 (2014).

21. Wu, F. *et al.* Onecut1 is essential for horizontal cell genesis and retinal integrity. *J. Neurosci.* **33**, 13053–65, 13065a (2013).

22. Gan, L. *et al.* POU domain factor Brn-3b is required for the development of a large set of retinal ganglion cells. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 3920–5 (1996).

23. Smith, A. M. *et al.* The transcription factor PU.1 is critical for viability and function of human brain microglia. *Glia* **61**, 929–42 (2013).

24. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

25. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

26. Gharahkhani, P. *et al.* Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat. Commun.* **12**, 1258 (2021).

27. Shahin, H. *et al.* Mutations in a novel isoform of TRIOBP that encodes a filamentous-actin binding protein are responsible for DFNB28 recessive nonsyndromic hearing loss. *Am. J. Hum. Genet.* **78**, 144–52 (2006).

28. Bonelli, R. *et al.* Identification of genetic factors influencing metabolic dysregulation and retinal support for MacTel, a retinal disorder. *Commun. Biol.* **4**, 274 (2021).

29. Thomas, E. D. *et al.* Multi-omic Analysis of Developing Human Retina and Organoids Reveals Cell-Specific Cis-Regulatory Elements and Mechanisms of Non-Coding Genetic Disease Risk. *bioRxiv* 2021.07.31.454254 (2021) doi:10.1101/2021.07.31.454254.

30. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–6 (2010).

31. Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).

32. Meng, W. *et al.* A genome-wide association study implicates that the TTC39C gene is associated with diabetic maculopathy with decreased visual acuity. *Ophthalmic Genet.* **40**, 252–258 (2019).

33. Shi, Y. *et al.* A genome-wide meta-analysis identifies two novel loci associated with high myopia in the Han Chinese population. *Hum. Mol. Genet.* **22**, 2325–33 (2013).

34. Park, S. J. H. *et al.* Function and Circuitry of VIP+ Interneurons in the Mouse Retina. *J. Neurosci.* **35**, 10685–700 (2015).

35. Yu, Y. *et al.* Common variants near FRK/COL10A1 and VEGFA are associated with advanced age-related macular degeneration. *Hum. Mol. Genet.* **20**, 3699–709 (2011).

36. Fritsche, L. G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat. Genet.* **45**, 433–9, 439e1-2 (2013).

37. Tideman, J. W. L. *et al.* Evaluation of Shared Genetic Susceptibility to High and Low Myopia and Hyperopia. *JAMA Ophthalmol.* **139**, 601–609 (2021).

38. Craig, J. E. *et al.* Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nat. Genet.* **52**, 160–166 (2020).

39. de Melo, J. *et al.* Injury-independent induction of reactive gliosis in retina by loss of function of the LIM homeodomain transcription factor Lhx2. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4657–62 (2012).

40. Hatakeyama, J., Tomita, K., Inoue, T. & Kageyama, R. Roles of homeobox and bHLH genes in specification of a retinal cell type. *Development* **128**, 1313–22 (2001).

41. Inoue, T. *et al.* Math3 and NeuroD regulate amacrine cell fate specification in the retina. *Development* **129**, 831–42 (2002).

42. Finkbeiner, C. *et al.* Single-cell ATAC-seq of fetal human retina and stem-cell-derived retinal organoids shows changing chromatin landscapes during cell fate acquisition. *Cell Rep.* **38**, 110294 (2022).

43. Fleckenstein, M. *et al.* Age-related macular degeneration. *Nat. Rev. Dis. Prim.* **7**, 31 (2021).

44. Duh, E. J., Sun, J. K. & Stitt, A. W. Diabetic retinopathy: current understanding, mechanisms, and treatment strategies. *JCI insight* **2**, (2017).

45. Weinreb, R. N., Aung, T. & Medeiros, F. A. The pathophysiology and treatment of glaucoma: a review. *JAMA* **311**, 1901–11 (2014).

46. Morgan, I. G., Ohno-Matsui, K. & Saw, S.-M. Myopia. *Lancet* **379**, 1739–1748 (2012).

47. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

48. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

49. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

50. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

51. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–40 (2012).

52. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

53. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic

features. *Bioinformatics* **26**, 841–2 (2010).

54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).

55. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

56. Myers, T. A., Chanock, S. J. & Machiela, M. J. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front. Genet.* **11**, 157 (2020).

57. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–11 (2001).

58. Tan, L. Determining the 3D genome structure of a single mammalian cell with Dip-C. *STAR Protoc.* **2**, 100622 (2021).

59. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).

60. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–80 (2014).

61. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101 (2016).

62. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2020).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

S.K.W. and H.Y.C. conceived the project and designed experiments. S.K.W., R.L., K.K., and C.L. performed experiments. S.K.W., S.N., A.P., A.P., and J.B.K. performed data analysis. A.K. and H.Y.C. supervised the work. S.K.W. and H.Y.C. wrote the manuscript with input from all authors.

COMPETING INTERESTS

H.Y.C. is a co-founder of Accent Therapeutics, Boundless Bio, Cartography Biosciences, and Circ Bio, and an advisor to 10X Genomics, Arsenal Biosciences, and Spring Discovery. A.K. is co-founder of RavelBio, a consulting fellow with Illumina, a member of the SAB of OpenTargets, PatchBio, SerImmune, and owns equity in DeepGenomics, Freenome and ImmunAI.

FIGURE LEGENDS

Figure 1. Transcriptional profiles from joint single-cell RNA- and ATAC-seq identify major cell types of the human retina.

a, Schematic of the human retina depicting the cell types analyzed in this study. **b**, Uniform manifold approximation and projection (UMAP) plot of the 51,645 human retinal cells detected by scRNA-seq after quality control filtering and removal of putative doublets. Eight postmortem retinas from four donors were profiled. A total of 22 clusters were resolved and assigned to 13 cell types. **c**, Frequency of different cell types in the human retina as determined by scRNA-seq. Numbers above each bar denote absolute counts out of 51,645. **d**, Dot plot visualizing the normalized RNA expression of selected marker genes by cell type. The color and size of each dot correspond to the average expression level and fraction of expressing cells, respectively.

Figure 2. Chromatin accessibility profiles from joint single-cell RNA- and ATAC-seq of the human retina reveal cell type-specific epigenetic landscapes.

a, Number of chromatin accessibility peaks for each cell type as determined by scATAC-seq. Peaks were required to be present in a least two pseudo-bulk ATAC replicates (n = 2 for Astrocyte and Microglia, n = 5 for all other cell types). **b**, Overlap of scATAC peaks with peaks from published human retina bulk ATAC-seq data. Overlapping was defined as peaks with any overlapping bases. **c**, Heatmap of scATAC marker peaks enriched in each cell type. Each column represents a marker peak. **d**, Sequencing tracks of chromatin accessibility near selected marker genes by cell type. Each track represents the aggregate scATAC signal of all cells from the given cell type normalized by the total number of reads in TSS regions. Genes in the sense direction (TSS on the left) are shown in red, while genes in the antisense direction (TSS on the right) are

shown in blue. Coordinates for each region: *PDE6A* (chr5:149924792-149964793), *GRIK1* (chr21:29905031–29955033), *RLBP1* (chr15:89201750–89241751), *GRM6* (chr5:178975297–179015298), *NIF3L1* (chr2:200874325–200914327), *ARR3* (chrX:70248304–70288305), *GAD2* (chr10:26186306–26246307), *ONECUT1* (chr15:52781076–52821078), *SLC6A9* (chr1:44005465–44035467), *NEFL* (chr8:24937109–24977110), *CALB2* (chr16:71323711–71368713), *PAX2* (chr10:100715602–100755603), *HLA-DRA* (chr6:32419841–32459842).

Figure 3. Motif analysis of accessible DNA regions in the human retina predicts cell type-specific transcription factors.

a, Heatmap of selected TF binding motifs enriched in each cell type. Darker colors indicate more significant enrichment. **b**, Footprinting analysis of selected TFs across cell types. Footprints were corrected for Tn5 insertion bias by dividing the footprinting signal by the Tn5 insertion signal.

Figure 4. Single-cell multiomics pinpoint the cellular targets of noncoding variants in eye diseases.

a, Overview of SNP selection for interrogating ocular disease GWAS. Index SNPs obtained from GWAS of each disease were subjected to LD expansion, and the resulting noncoding SNPs intersected with scATAC peaks. **b**, Percentage of LD expanded SNPs from each disease that overlapped with chromatin accessibility peaks for each cell type. **c**, Schematic of promoter co-accessibility and predicted target gene analyses. Co-accessible was defined as scATAC peaks whose accessibility showed a correlation score >0.3. Predicted target genes were defined as genes whose RNA expression showed a correlation score >0.3 relative to the accessibility of the

tested scATAC peak. **d**, Number of scATAC peaks co-accessible with each promoter peak. **e**, Number of predicted target genes for each scATAC peak. **f**, Sequencing tracks of chromatin accessibility near rs4821699 (chr22:37719685) and rs17421627 (chr5:88551768). Genes in the sense and antisense directions are shown in red and blue, respectively. The location of each SNP is depicted by a vertical gray line. Gray arcs indicate predicted target genes for the scATAC peak containing the SNP of interest.

Figure 5. Integration of single-cell multiome with HiChIP and eQTL data prioritizes functional noncoding polymorphisms in the human retina.

a, Schematic of H3K27ac HiChIP and eQTL analyses used to prioritize SNPs. **b**, Overlap of HiChIP loop anchors (n = 2 biological replicates) with scATAC peaks. **c**, Percentage of SNPs in scATAC peaks for each disease with available retina eQTL data. **d**, Sequencing tracks of chromatin accessibility near rs9966620 (chr18:24100771), rs2730260 (chr7:159054238), and rs66475830 (chr6:116087639). Genes in the sense and antisense directions are shown in red and blue, respectively. The location of each SNP is depicted by a vertical gray line. Gray arcs indicate predicted target genes for the scATAC peak containing the SNP of interest. The black arc overlapping with rs9966620 indicates a H3K27ac HiChIP loop with the region encompassed by the opposite anchor highlighted in purple. **e**, Significance of SNP-gene associations for rs2730260 or rs66475830 and their nearby genes as determined by retina eQTL analysis. Adjusted P values for each gene were calculated by multiplying the nominal P value listed in the EyeGEx database by the number of SNP-gene pairs tested for that SNP.

Figure 6. Integration of single-cell multiome with base-resolution deep learning nominates functional mechanisms for disease-associated SNPs.

a, Schematic of the CNN-based deep learning pipeline. **b**, Percentage of noncoding index SNPs (n = 1,284), LD expanded SNPs (n = 7,034), LD expanded SNPs in scATAC peaks (n = 1,152), randomly selected GC-matched SNPs (n = 9,984), and randomly selected SNPs in scATAC peaks (n = 1,160) that were categorized as high effect. **c**, Top: Predicted per-base accessibility for rs1532278 (chr8:27608798) and rs1874459 (chr16:65041801) in Müller glia and rod bipolar cells, respectively, as determined by deep learning models. A 100-bp window depicts the importance of each base to predicted accessibility at the SNP, and a 1,000-bp window depicts predicted per-base counts for the reference (blue) and alternate (orange) alleles. SNP bases are highlighted in purple. For rs1874459, similar changes in accessibility were predicted for OFF-cone bipolar, ON-cone bipolar, gly-amacrine, and AII-amacrine cells. Bottom: Sequencing tracks of chromatin accessibility near rs1532278 and rs1874459. Genes in the sense and antisense directions are shown in red and blue, respectively. The location of each SNP is depicted by a vertical gray line. Gray arcs indicate predicted target genes for the scATAC peak containing the SNP of interest. **d,f**, Significance of SNP-gene associations for rs1532278 (d) or rs1874459 (f) and their nearby genes as determined by retina eQTL analysis. Adjusted *P* values for each gene were calculated by multiplying the nominal *P* value listed in the EyeGEx database by the number of SNP-gene pairs tested for that SNP. **e**, Dot plot visualizing the normalized RNA expression of 40 different homeodomain TFs in Müller glia. The selected TFs correspond to the 40 homeodomain factors whose binding motifs were most significantly enriched in Müller glia as determined by motif analysis (Supplementary Data 4). **g**, Dot plot visualizing the normalized RNA expression of neuroD and neurogenin family members by cell type.

Extended Data Figure 1. Single-cell RNA- and ATAC-seq quality control metrics.

a-f, Violin plots depicting the number of detected genes per nucleus (a), number of detected RNA transcripts per nucleus (b), percentage of mitochondrial reads (c), percentage of ribosomal reads (d), TSS enrichment (e), and number of detected fragments per nucleus (f) by retina in the final dataset. Boxes depict the 25th percentile, median, and 75th percentile of each metric.

Extended Data Figure 2. Single-cell RNA-seq cluster assignments.

a, First iteration uniform manifold approximation and projection (UMAP) plot of the scRNA-seq dataset after quality control filtering, automated removal of doublets, and exclusion of clusters with no detected marker genes. Five clusters comprised of putative doublets were subsequently removed and the dataset re-clustered. **b**, Second iteration UMAP of the scRNA-seq dataset. One cluster comprised of putative doublets was subsequently removed and the dataset re-clustered. **c**, Final iteration UMAP of the scRNA-seq dataset. Clusters representing subpopulations of the same cell type were grouped together for downstream analyses.

Extended Data Figure 3. Characterization of chromatin accessibility peaks and linkage disequilibrium expanded SNPs.

a, Number of cell types exhibiting each of the 620,386 scATAC peaks. **b**, Visual depiction of LD expansion for the 30 index SNPs on chromosome 1 associated with AMD. Each vertical black line represents a SNP. **c**, Percentage of noncoding index and LD expanded SNPs from each disease that overlapped with at least one scATAC peak. **d**, Number of cell types exhibiting each of the 1,152 scATAC peaks that overlapped with a LD expanded SNP. **e**, Comparison of

predicted target genes with the nearest gene for each chromatin accessibility peak with at least one predicted target gene. Same was defined as any of the predicted target genes being the nearest gene.

Extended Data Figure 4. H3K27ac HiChIP predicts enhancer-promoter interactions in the human retina.

a, HiChIP interaction types by sample. Cis refers to interactions on the same chromosome, while trans refers to interactions spanning separate chromosomes. **b**, HiChIP interaction maps at whole genome, 250-kb, and 5-kb resolution. Sample shown is LGS1OD. Signal was normalized to the square root of coverage. Numbers above the interaction maps indicate maximum signal in each matrix. **c**, Sequencing tracks of cell type chromatin accessibility and H3K27ac HiChIP loops overlapping with the TSS of selected marker genes. Genes in the sense and antisense directions are shown in red and blue, respectively. Regions encompassed by the loop anchors are highlighted in purple.

Extended Data Figure 5. Integration of single-cell multiome with HiChIP and eQTL data refines SNP-target gene predictions.

a, Sequencing tracks of chromatin accessibility near rs77272443 (chr1:112894884) and rs4102217 (chr11:65496424). Genes in the sense and antisense directions are shown in red and blue, respectively. The location of each SNP is depicted by a vertical gray line. Gray arcs indicate predicted target genes for the scATAC peak containing the SNP of interest. Black arcs overlapping with SNPs indicate H3K27ac HiChIP loops with the regions encompassed by the opposite anchors highlighted in purple. **b**, Significance of SNP-gene associations for rs77272443

or rs4102217 and their 20 nearest genes as determined by retina eQTL analysis. Adjusted *P* values for each gene were calculated by multiplying the nominal *P* value listed in the EyeGEx database by the number of SNP-gene pairs tested for that SNP.

Extended Data Figure 6. BPNet model performance.

a,b, Correlation scores between predicted and actual observed log counts in only peak (a) or both peak and non-peak (b) regions on chromosome sequences that were withheld during cell type-specific model training. Each color represents one of five model folds.

Extended Data Figure 7. Summary of SNP prioritization.

a-e, Total number of prioritization criteria met by LD expanded SNPs in loci associated with AMD (a), glaucoma (b), DR (c), myopia (d), and MacTel (e). Prioritization criteria were 1) co-accessibility with a promoter peak, 2) accessibility correlated with expression of a nearby gene, 3) linkage to a gene by H3K27ac HiChIP, 4) significant association with a gene by retina eQTL data, and 5) high effect designation by deep learning.

a

Rod
Cone
Horizontal
ON-cone bipolar
Rod bipolar
GABA-amacrine
Retinal ganglion cell

Microglia
Müller glia
OFF-cone bipolar
Gly-amacrine
AII-amacrine
Astrocyte

b

- Rod
- OFF-cone bipolar (5 clusters)
- Müller glia (2 clusters)
- ON-cone bipolar (4 clusters)
- Rod bipolar
- Cone
- GABA-amacrine (2 clusters)
- Horizontal
- Gly-amacrine
- Retinal ganglion cell
- AII-amacrine
- Astrocyte
- Microglia

UMAP_2 / UMAP_1

c

n = 51,645 cells

% of cells

32,059
3,936
3,670
2,617
2,324
2,223
1,543
1,135
730
493
451
232
232

Rod, OFF-cone bipolar, Müller glia, ON-cone bipolar, Rod bipolar, Cone, GABA-amacrine, Horizontal, Gly-amacrine, Retinal ganglion cell, AII-amacrine, Astrocyte, Microglia

d

Percent expressed
0, 25, 50, 75

Average expression
2.5, 2.0, 1.5, 1.0, 0.5, 0.0, -0.5

Cell type (y-axis): Rod, OFF-cone bipolar, Müller glia, ON-cone bipolar, Rod bipolar, Cone, GABA-amacrine, Horizontal, Gly-amacrine, Retinal ganglion cell, AII-amacrine, Astrocyte, Microglia

Genes (x-axis): PDE6A, NR2E3, GRIK1, RLBP1, GLUL, GRM6, ISL1, PRKCA, NIF3L1, ARR3, OPN1LW, GAD1, GAD2, ONECUT1, ONECUT2, SLC6A9, NEFL, NEFM, GJD2, CALB2, GFAP, PAX2, HLA-DRA, C1QA

a # of peaks (x10³) by cell type, colored by Distal, Exonic, Intronic, Promoter: Rod, OFF-cone bipolar, Müller glia, ON-cone bipolar, Rod bipolar, Cone, GABA-amacrine, Horizontal, Gly-amacrine, Retinal ganglion cell, AII-amacrine, Astrocyte, Microglia, Union

b # of peaks (x10³): Bulk ATAC, scATAC — Overlapping, Unique

c 197,826 peaks; Column Z−Scores −2.0 to 2.0; rows: Rod, OFF-cone bipolar, Müller glia, ON-cone bipolar, Rod bipolar, Cone, GABA-amacrine, Horizontal, Gly-amacrine, Retinal ganglion cell, AII-amacrine, Astrocyte, Microglia

d Gene tracks: *PDE6A*, *GRIK1*, *RLBP1*, *GRM6*, *NIF3L1*, *ARR3*, *GAD2*, *ONECUT1*, *SLC6A9*, *NEFL*, *CALB2*, *PAX2*, *HLA-DRA* across cell types: Rod, OFF-cone bipolar, Müller glia, ON-cone bipolar, Rod bipolar, Cone, GABA-amacrine, Horizontal, Gly-amacrine, Retinal ganglion cell, AII-amacrine, Astrocyte, Microglia, Genes

a

	TCF3	TCF4	ZEB1	NEUROD1	FOS	JUND	LHX2	LHX9	DBP	HLF	TEF	POU2F1	OTX2	MEF2C	NFATC3	GATA3	RARA	MEIS1	MEIS2	RFX2	RFX3	RFX5	ONECUT1	ONECUT2	ONECUT3	CUX1	CUX2	NHLH1	PKNOX2	EBF1	POU4F1	POU4F2	POU4F3	RORA	NR4A1	OLIG1	NEUROG2	TFAP4	BHLHE22	EGR3	ZNF263	SP1	SP3	SPI1	ETV6	ELF1
Rod																																														
OFF-cone bipolar																																														
Müller glia																																														
ON-cone bipolar																																														
Rod bipolar																																														
Cone																																														
GABA-amacrine																																														
Horizontal																																														
Gly-amacrine																																														
Retinal ganglion cell																																														
AII-amacrine																																														
Astrocyte																																														
Microglia																																														

$-\log_{10}(\text{adjusted } P)$

0   50   100   150   200

b

OTX2, LHX2, ONECUT1, SPI1

Tn5 bias-corrected normalized insertions

Distance to motif center (bp)

— Rod
— OFF-cone bipolar
— Müller glia
— ON-cone bipolar
— Rod bipolar
— Cone
— GABA-amacrine
— Horizontal
— Gly-amacrine
— Retinal ganglion cell
— AII-amacrine
— Astrocyte
— Microglia

a

	AMD	Glaucoma	DR	Myopia	MacTel
# index SNPs	172	361	130	653	22
# noncoding SNPs after LD expansion	708	2,074	864	3,356	81

7,034 unique SNPs

Intersection with scATAC peaks

| 124 | 377 | 59 | 581 | 13 |

b

% of LD expanded SNPs overlapping peaks

- Rod
- OFF-cone bipolar
- Müller glia
- ON-cone bipolar
- Rod bipolar
- Cone
- GABA-amacrine
- Horizontal
- Gly-amacrine
- Retinal ganglion cell
- AII-amacrine
- Astrocyte
- Microglia

AMD / Glaucoma / DR / Myopia / MacTel

c Co-accessible w/ promoter peak

Predicted target genes

d n = 39,552 promoter peaks

% of promoter peaks vs # of co-accessible peaks (0, 1, 2-5, 6-10, 11-20, >20)

e n = 620,386 peaks

% of peaks vs # of predicted target genes (0, 1, 2, 3, 4, >4)

f

chr22 rs4821699 (Glaucoma)   chr5 rs17421627 (MacTel)

Rod / OFF-cone bipolar / Müller glia / ON-cone bipolar / Rod bipolar / Cone / GABA-amacrine / Horizontal / Gly-amacrine / Retinal ganglion cell / AII-amacrine / Astrocyte / Microglia

Genes: *TRIOBP*   *LINC00461*

37,700,000   37,750,000   88,525,000   88,575,000

a H3K27ac HiChIP + Retina eQTLs from EyeGEx

b n = 9,670 H3K27ac HiChIP loops

% of loops

- No peak in either anchor
- Peak in one anchor
- Peak in both anchors

c % of accessible SNPs with retina eQTL data

- AMD
- Glaucoma
- DR
- Myopia
- MacTel

d

chr18 — rs9966620 (DR)
chr7 — rs2730260 (Myopia)
chr6 — rs66475830 (AMD)

Rod
OFF-cone bipolar
Müller glia
ON-cone bipolar
Rod bipolar
Cone
GABA-amacrine
Horizontal
Gly-amacrine
Retinal ganglion cell
AII-amacrine
Astrocyte
Microglia

Genes

TTC39C *OSBPL1A*
TTC39C-AS1

WDR60 *VIPR2*
LINC00689

COL10A1 *TSPYL4*
FRK *DSE*
NT5DC1 *TSPYL1*

H3K27ac HiChIP loops

24,050,000   24,150,000
159,000,000   159,100,000
116,000,000   116,200,000

e

rs2730260
$-\log_{10}(\text{adjusted } P)$
ESYT2, *LINC00689*, *NCAPG2*, *PTPRN2*, *VIPR2*, *WDR60*

rs66475830
$-\log_{10}(\text{adjusted } P)$
COL10A1, DSE, FRK, KPNA5, NT5DC1, RSPH4A, RWDD1, TSPYL1, TSPYL4, ZUP1

a Cell-type specific ATAC-seq fragments → CNN-based deep learning → Predicted per-base accessibility by cell type

b % high effect

c chr8 rs1532278 (Myopia) [-homeodomain]   chr16 rs1874459 (Glaucoma) [+bHLH domain]

d rs1532278

e Müller glia (homeodomain TFs)

f rs1874459

g

a, **b**, **c**

Legend (panel a):
- Rod
- OFF-cone bipolar (1)
- OFF-cone bipolar (2)
- OFF-cone bipolar (3)
- OFF-cone bipolar (4)
- OFF-cone bipolar (5)
- Müller glia (1)
- Müller glia (2)
- ON-cone bipolar (1)
- ON-cone bipolar (2)
- ON-cone bipolar (3)
- ON-cone bipolar (4)
- Rod bipolar
- Cone
- GABA-amacrine (1)
- GABA-amacrine (2)
- Horizontal
- Gly-amacrine
- Retinal ganglion cell
- AII-amacrine
- Astrocyte
- Microglia
- Müller/Bipolar doublets
- Müller/Horizontal doublets
- ON/OFF bipolar doublets
- Müller/GABA-amacrine doublets
- Müller/Cone doublets

Legend (panel b):
- Rod
- OFF-cone bipolar (1)
- OFF-cone bipolar (2)
- OFF-cone bipolar (3)
- OFF-cone bipolar (4)
- OFF-cone bipolar (5)
- Müller glia (1)
- Müller glia (2)
- ON-cone bipolar (1)
- ON-cone bipolar (2)
- ON-cone bipolar (3)
- ON-cone bipolar (4)
- Rod bipolar
- Cone
- GABA-amacrine (1)
- GABA-amacrine (2)
- Horizontal
- Gly-amacrine (2)
- Gly-amacrine (2)
- Retinal ganglion cell
- AII-amacrine
- Astrocyte
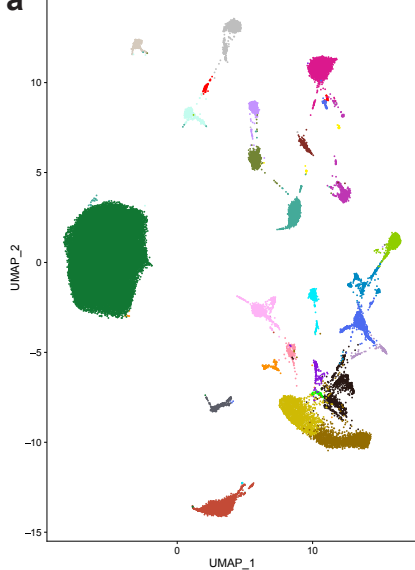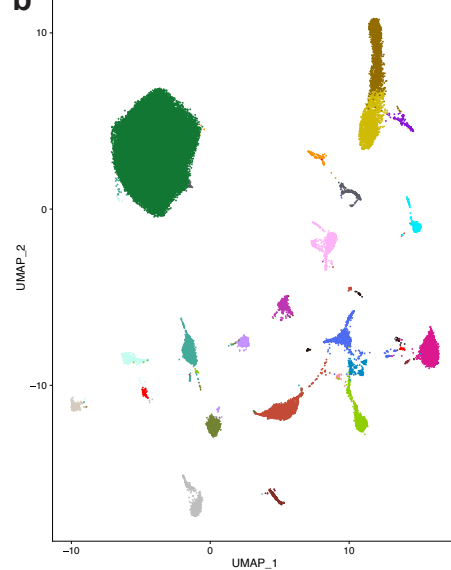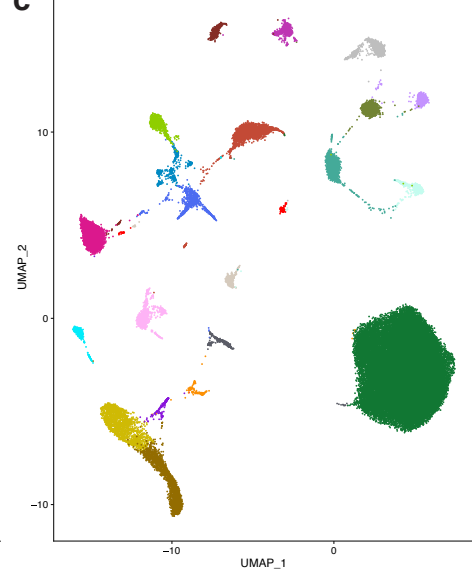- Microglia
- ON/OFF bipolar doublets

Legend (panel c):
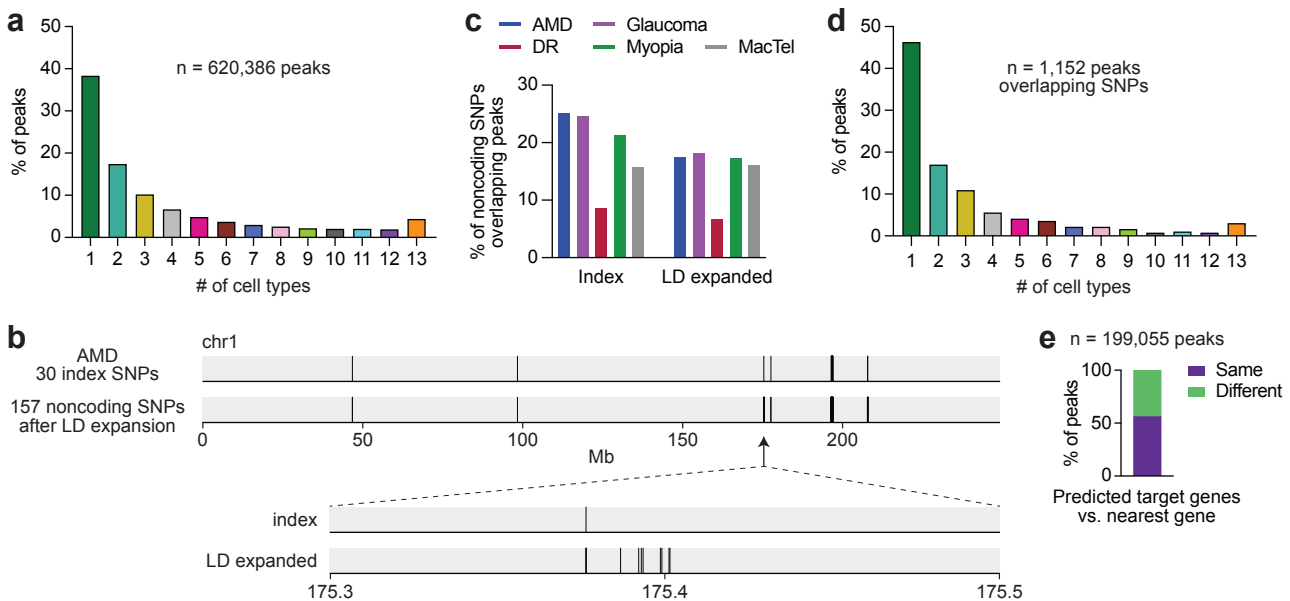- Rod
- OFF-cone bipolar (1)
- OFF-cone bipolar (2)
- OFF-cone bipolar (3)
- OFF-cone bipolar (4)
- OFF-cone bipolar (5)
- Müller glia (1)
- Müller glia (2)
- ON-cone bipolar (1)
- ON-cone bipolar (2)
- ON-cone bipolar (3)
- ON-cone bipolar (4)
- Rod bipolar
- Cone
- GABA-amacrine (1)
- GABA-amacrine (2)
- Horizontal
- Gly-amacrine
- Retinal ganglion cell
- AII-amacrine
- Astrocyte
- Microglia

a % of peaks vs. # of cell types. n = 620,386 peaks

b chr1. AMD 30 index SNPs; 157 noncoding SNPs after LD expansion; Mb; index; LD expanded; 175.3, 175.4, 175.5

c % of noncoding SNPs overlapping peaks. AMD, DR, Glaucoma, Myopia, MacTel. Index, LD expanded

d % of peaks vs. # of cell types. n = 1,152 peaks overlapping SNPs

e n = 199,055 peaks. % of peaks. Same, Different. Predicted target genes vs. nearest gene

a

Interaction type
- Cis short (<20 kb)
- Cis long (>20 kb)
- Trans

b

= 8962   250-kb resolution   = 170   5-kb resolution   = 3

chr1-22,X,Y

chr1-22,X,Y

0 Mb ... 145.13 Mb   chr8   0 Mb ... 145.13 Mb

25 Mb ... 28.53 Mb

CLU

c

chr5   Rod   chr8   Müller glia   chr5   Rod bipolar

Genes

PDE6A   *CLU*   *LINC02106*   *ISL1*

SLC26A2   *MIR6843*   *CCDC25*   *ISL1-DT*

H3K27ac
HiChIP loops

149,900,000   149,950,000   150,000,000

27,600,000   27,690,000   27,780,000

51,000,000   51,175,000   51,350,000

a

chr1   rs77272443 (Myopia)

chr11   rs4102217 (Glaucoma)

Rod
OFF-cone bipolar
Müller glia
ON-cone bipolar
Rod bipolar
Cone
GABA-amacrine
Horizontal
Gly-amacrine
Retinal ganglion cell
AII-amacrine
Astrocyte
Microglia

Genes

LINC01356   *SLC16A1-AS1*   *NEAT1*   *MALAT1*   *EHBP1L1*
FAM19A3   *SLC16A1*   *FRMD8*   *SCYL1*   *FAM89B*
MIR612   *SSSCA1*
AL357055.3   *LTBP3*
SSSCA1-AS1   *SIPA1*

H3K27ac HiChIP loops

112,800,000   113,000,000   65,450,000   65,550,000

b

rs77272443

$-\log_{10}(\text{adjusted } P)$

AL133517.1
AL357055.3
AL390729.1
AL603832.3
CAPZA1
CTTNBP2NL
FAM19A3
LINC01356
LINC01750
LRIG2
MAGI3
MOV10
PHTF1
PPM1J
RHOC
RSBN1
SLC16A1
SLC16A1-AS1
ST7L
WNT2B

rs4102217

$-\log_{10}(\text{adjusted } P)$

AP000769.1
CDC42EP2
DPF2
EHBP1L1
FAM89B
FRMD8
KAT5
LTBP3
MALAT1
MAP3K11
NEAT1
PCNX3
POLA2
RELA
SCYL1
SIPA1
SLC25A45
SSSCA1
SSSCA1-AS1
TIGD3

a Peaks only

b Peaks and non-peaks

a AMD

# of SNPs

Not in peak: 584

# of prioritization criteria met
- 0: 44
- 1: 38
- 2: 25
- 3: 11
- 4: 6
- 5: 0

b Glaucoma

# of SNPs

Not in peak: 1,697

# of prioritization criteria met
- 0: 86
- 1: 150
- 2: 101
- 3: 37
- 4: 3
- 5: 0

c DR

# of SNPs

Not in peak: 805

# of prioritization criteria met
- 0: 25
- 1: 23
- 2: 7
- 3: 3
- 4: 1
- 5: 0

d Myopia

# of SNPs

Not in peak: 2,775

# of prioritization criteria met
- 0: 163
- 1: 201
- 2: 133
- 3: 61
- 4: 22
- 5: 1

e MacTel

# of SNPs

Not in peak: 68

# of prioritization criteria met
- 0: 2
- 1: 6
- 2: 4
- 3: 1
- 4: 0
- 5: 0