Mapping the *cis*-regulatory architecture of the human retina reveals noncoding genetic variation in disease

Timothy J. Cherry[a,b,c,d,1], Marty G. Yang[e,f], David A. Harmin[e], Peter Tao[e], Andrew E. Timms[d], Miriam Bauwens[g], Rando Allikmets[h,i], Evan M. Jones[j,k], Rui Chen[j,k], Elfride De Baere[g], and Michael E. Greenberg[e,1]

[a]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98101; [b]Department of Ophthalmology, University of Washington School of Medicine, Seattle, WA 98101; [c]Department of Biological Structure, University of Washington School of Medicine, Seattle, WA 98101; [d]Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA 98101; [e]Department of Neurobiology, Harvard Medical School, Boston, MA 02115; [f]Program in Neuroscience, Harvard Medical School, Boston, MA 02115; [g]Center for Medical Genetics, Ghent University and Ghent University Hospital, 9000 Ghent, Belgium; [h]Department of Ophthalmology, Columbia University, New York, NY 10032; [i]Department of Pathology and Cell Biology, Columbia University, New York, NY 10032; [j]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030-3411; and [k]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030-3411

Contributed by Michael E. Greenberg, February 27, 2020 (sent for review December 23, 2019; reviewed by Seth Blackshaw and Joseph C. Corbo)

The interplay of transcription factors and *cis*-regulatory elements (CREs) orchestrates the dynamic and diverse genetic programs that assemble the human central nervous system (CNS) during development and maintain its function throughout life. Genetic variation within CREs plays a central role in phenotypic variation in complex traits including the risk of developing disease. We took advantage of the retina, a well-characterized region of the CNS known to be affected by pathogenic variants in CREs, to establish a roadmap for characterizing regulatory variation in the human CNS. This comprehensive analysis of tissue-specific regulatory elements, transcription factor binding, and gene expression programs in three regions of the human visual system (retina, macula, and retinal pigment epithelium/choroid) reveals features of regulatory element evolution that shape tissue-specific gene expression programs and defines regulatory elements with the potential to contribute to Mendelian and complex disorders of human vision.

enhancer | *cis*-regulatory element | human | retina | noncoding

The human central nervous system (CNS) is composed of diverse tissues and cell types, whose unique forms and functions arise from their distinct programs of gene expression. The precise timing, levels, and cell type specificity of gene expression are controlled by cohorts of *cis*-regulatory elements (CREs), including promoters and distal enhancers. These short DNA sequences (100 to 500 bp) consist of binding sites for sequence-specific transcription factors (TFs) and function coordinately to recruit transcriptional machinery to specific sites in the genome (1). Genetic variation within the sequence of CREs, between individuals within a population or between species, can lead to changes in TF binding and thus changes in the expression of linked gene(s). These regulatory variants are now appreciated to play a central role in the evolution of species-specific traits and in phenotypic variation between individuals including disease (2–6). A major obstacle toward understanding normal and pathological consequences of regulatory variation in human tissues is that many cell type- and tissue-specific CREs have not been comprehensively mapped in the relevant human tissues. Thus, their functions in development and tissue homeostasis remain incompletely understood. Accordingly, the identification and characterization of CREs have become a major focus of genetics in the early 21st century (7).

The retina is a classic model for CNS development, function, and gene regulation because of its highly stereotyped cellular composition and circuit organization. This reduced cellular diversity compared to other CNS regions has recently made the retina a powerful system for characterizing the mechanisms by which TFs function at CREs to control the specification and differentiation of retinal cell types (8–11). The retina consists of six major cell classes that are specified from a common pool of progenitors during development through the cooperation of broadly expressed and lineage-defining TFs. The largest fraction of cells in the human retina are rod photoreceptor cells, which mediate the earliest steps in vision. The retinal pigment epithelium (RPE) is directly apposed to mammalian photoreceptor cells and is required for photoreceptor cell survival and function. The retina and RPE share an embryonic origin in the early optic vesicle; however, as mature tissues, they are functionally and molecularly distinct. Together, the retina, the RPE, and the macula, a specialized central region of the human retina that mediates high-acuity vision, are commonly affected in human visual disorders. It is therefore important to understand the genetic regulation of these tissues to serve as a model of CNS function and disease.

TFs that specify the cell types of the retina and RPE are highly conserved across species. Mutations in TFs that disrupt gene expression in specific cell types often result in similar phenotypes in mice and humans including congenital blindness, photoreceptor degeneration, ocular malformations, and glaucoma (12–15). It is unclear, however, to what extent the CREs that recruit these TFs are conserved between species. Indeed, the evolutionary conservation of CREs is significantly lower than that of protein coding genes (3). An increasing number of studies show that mutations in

Significance

Regulation of gene expression is critical for all complex biological processes including vision. Mutations within *cis*-regulatory elements (CREs) can disrupt gene expression and contribute to human diseases. However, it is challenging to screen for mutations within CREs that affect vision because the genomic locations and mechanisms of action of these elements are largely unknown. Here, we take advantage of advances in epigenomic and transcriptomic profiling techniques to identify and characterize CREs in three regions of the human eye commonly affected in visual disorders. Our data identify shared and unique CREs in each region, reveal the combinatorial binding of TFs, and provide a resource for understanding the role of noncoding genetic variation in visual disorders.

Author contributions: T.J.C., M.G.Y., and M.E.G. designed research; T.J.C., M.G.Y., and P.T. performed research; T.J.C., M.G.Y., D.A.H., M.B., R.A., E.M.J., R.C., and E.D.B. contributed new reagents/analytic tools; T.J.C., M.G.Y., D.A.H., P.T., A.E.T., and M.E.G. analyzed data; and T.J.C., M.G.Y., and M.E.G. wrote the paper.

Reviewers: S.B., Johns Hopkins University School of Medicine; and J.C.C., Washington University School of Medicine.

The authors declare no competing interest.

Data deposition: All high-throughput sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE137311).

[1]To whom correspondence may be addressed. Email: timothy.cherry@seattlechildrens.org or michael_greenberg@hms.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922501117/-/DCSupplemental.

First published April 7, 2020.

specific CREs cause retinal disorders (16–18). It is therefore critical to map and characterize the CREs that regulate essential gene expression in human eye tissues to refine the genomic search-space for additional human disease-causing noncoding mutations. These analyses will also facilitate discovery of the mechanisms by which TFs selectively bind and activate cell type-specific CREs to regulate the expression of genes essential for proper retinal development and function.

In this study, we used an integrated epigenomic approach to identify and functionally characterize CREs in three tissue regions that are essential for human vision: the retina, macula, and RPE/choroid. By comparing the epigenetic landscape of these tissues, we found that tissue-specific CREs are regulated primarily at the level of differential chromatin accessibility and through distinct cohorts of TFs. We assessed the genomic binding of five TFs that are essential for retinal function and found that combinations of TFs are essential rather than being redundant for CRE function. To investigate the role of TF/CRE interactions in individual retinal cell types, we profiled gene expression in >4,000 single human retinal nuclei and identified candidate cell type-specific CREs. Comparison of human and mouse retinal CREs demonstrated functional conservation of CREs at essential retinal genes and highlighted species-specific differences that alter gene expression. Last, we were able to show that CRE identification can facilitate the discovery and interpretation of disease-associated regulatory variation in CREs. These findings provide recent insights into the mechanisms by which regulatory variation shapes the development of complex neural tissues, species-specific traits, and disease outcomes affecting human vision.

Results

Identification of Active CREs in Human Visual Tissues. To identify the specific CREs in the genome that control genes necessary for human visual function, we first sought to identify shared and tissue-specific CREs in adult human retina, macula, and RPE/choroid based on epigenetic features of enhancers and promoters (Fig. 1 *A* and *B*). Candidate active CREs in each tissue type can be defined by the following: 1) accessible chromatin as a result of TF binding, 2) enrichment for histone modifications associated with active enhancers or promoters, and 3) association with active gene expression (Fig. 1*B*). These principal features of CREs were profiled and examined for subsequent analyses in this study (Fig. 1*C*). Chromatin accessibility was determined using assay for transposase-accessible chromatin sequencing (ATAC-seq) (19) on nuclei purified from adult postmortem unfixed human retina and macula (Fig. 1*D* and *SI Appendix*, Fig. S1*A*). DNase-seq datasets were used to identify chromatin accessibility in RPE (20) (Fig. 1*D*). A read-normalized comparison of ATAC-seq vs. DNase-seq in mouse retinal tissue demonstrated correlation ($r^2 = 0.69$) and suggested these assays are generally comparable for purposes of identifying regions of DNA accessibility (*SI Appendix*, Fig. S1 *B* and *C*). To distinguish active CREs from other accessible regions, we colocalized accessible regions with the histone mark consistent with active enhancers and promoters, acetylated lysine 27 on histone 3 (H3K27ac), using chromatin immunoprecipitation sequencing (ChIP-seq) (21, 22). Altogether, we identified 21,368, 19,933, and 12,212 regions that were enriched for DNA accessibility and H3K27ac in the retina, macula, and RPE/choroid, respectively (Fig. 1*D* and Dataset S2). To facilitate the use of these data in basic and disease research, we have assembled them, and additional data described below, into a searchable track hub via the University of California, Santa Cruz (UCSC) genome browser (https://tinyurl.com/CherryLab-EyeBrowser) (*SI Appendix*).

Active CREs engage transcriptional machinery to promote target gene expression. To confirm that these identified regions are associated with active gene expression in each tissue, we performed RNA sequencing (RNA-seq) on adult human retina, macula, and RPE/choroid. RNA-seq was performed on total RNA extracted from cell nuclei (total RNA Nuc-seq), to avoid cross contamination of RNA between the intercalated cells of the retina and RPE/choroid, and to enrich for enhancer RNAs (eRNAs), a marker of active enhancers (23). We then compared the level of proximal gene expression to the level of H3K27ac at putative enhancers and promoters. We found that levels of H3K27ac are correlated with levels of proximal gene expression (Fig. 1 *D* and *E* and *SI Appendix*, Fig. S1*D*). Highly expressed genes associated with active CREs in each tissue include the known disease genes *RP1*, *ABCA4*, and *RPE65* in retina, macula, and RPE/choroid, respectively (Fig. 1*F*). At each of these gene loci, proximal promoters and one or more distal enhancers were identified based on the co-occurrence of DNA accessibility, H3K27ac signal, and, when present, eRNAs. The correlation of each of these assays was quantified between biological replicates and across tissues to assess reproducibility and tissue specificity (*SI Appendix*, Fig. S1*A*).

Active enhancers and promoters in each tissue contribute to the unique programs of gene expression and biological function of each tissue. To determine the distinct biological functions associated with active enhancers and promoters in each tissue, we performed genome regions enrichment of annotations (GREAT) analysis (24). We found CREs in the retina and macula to be significantly associated with regulation of rhodopsin signaling and the cellular response to light, consistent with the abundance of photoreceptor cells in both tissues (Fig. 1*G*). Active regulatory elements in the RPE/choroid, however, were associated with regulation of ion transporter activity, an essential function of RPE to compensate for light-dependent changes in photoreceptor capacitance (25, 26). This analysis also uncovered association of active CREs with distinct pathologies in each tissue including night blindness, abnormal retinal vasculature, and abnormalities of the vitreous in retina, macula, and RPE, respectively.

To validate the transcriptional activity of CREs identified using this integrated epigenomic approach, we cloned candidate CREs from the human *ABCA4* Stargardt disease gene locus (*SI Appendix*, Fig. S1*E*) into a luciferase reporter construct and electroporated these constructs individually into the postnatal day 0 (P0) mouse retina and harvested at P11 as *ABCA4*-expressing photoreceptor cells are maturing. Although this approach is not sufficient to determine the expected photoreceptor cell-specific activity of these enhancers, we did observe that the majority of these identified CREs (six of eight) increased the transcriptional output of the reporter construct when compared to the empty vector control, suggesting that these elements have intrinsic CRE activity (*SI Appendix*, Fig. S1*F*). Taken together, these integrative epigenomic and transcriptomic analyses demonstrate that active regulatory elements contribute to unique biological processes in the human retina, macula, and RPE/choroid. Disruption of these regulatory elements by genetic variation could have tissue-specific consequences resulting in visual disorders. The results of these analyses have now defined the sites in the genome in which to look for *cis*-regulatory variants implicated in disease.

Tissue-Specific Gene Expression Is Shaped by DNA Accessibility of CREs. The retina, macula, and RPE share a common embryonic origin in the nascent optic vesicle; however, each tissue performs a distinct function in adult vision. Most CRE accessibility is shared between retina and macula (68% and 69%, reciprocally), likely due to the shared cell types that make up these tissues (Fig. 2*A*). In contrast, only a fraction of their DNA-accessible sites is shared with RPE/choroid (23% and 17% reciprocally). This suggests that retina vs. RPE/choroid CREs are generally maintained by differential DNA accessibility. Among CREs that are accessible in all tissues, most also share H3K27ac enrichment (65%) (Fig. 2*B*). This shows that while tissue-specific CRE activation does occur, differential accessibility is the dominant pattern of tissue-specific CREs. Examples of both modes of regulation can be observed at the human disease gene loci *PRPH2*, *RPE65*, and *ABCA4* (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S2*A*).

Intriguingly, genes that are expressed in both retina and RPE can also employ tissue-specific CREs giving each tissue "private" control over shared genes. For example, we found that ABCA4, the mutated gene in autosomal recessive Stargardt disease (STGD1)

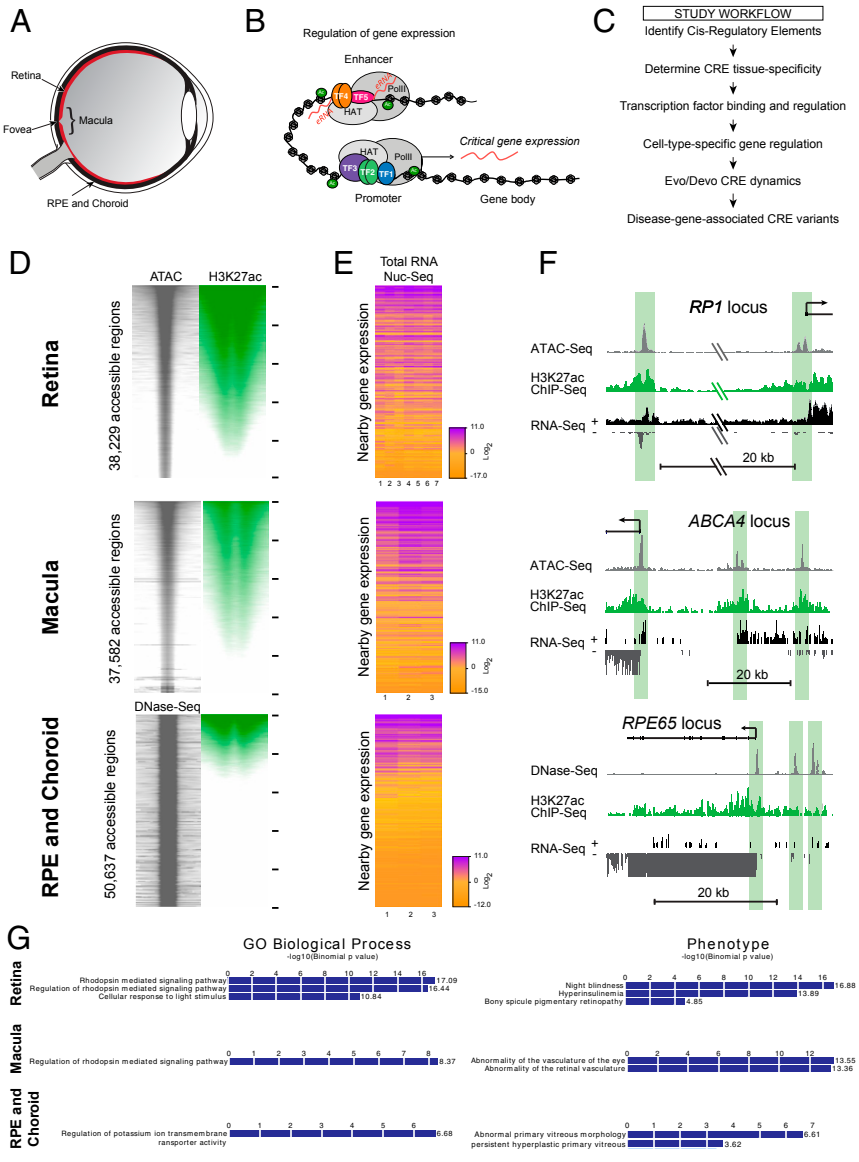

Fig. 1. Identification and characterization of active *cis*-regulatory elements (CREs) in human retina, macula, and RPE. (*A*) Schematic cross-section of the human eye with tissues commonly affected in inherited retinal diseases labeled. (*B*) Schematic of *cis*-regulatory control of gene expression. Transcription factors (TF1–5) bind in combination to promoters and enhancers to recruit cofactors including histone acetyl transferases (HATs) and basal transcriptional machinery such as RNA polymerase (PolII). HATs acetylate lysine residues on histones, including H3K27ac (Ac). PolII can induce transcription of enhancer RNAs (eRNAs). Both H3K27ac and eRNAs are associated with active CREs. (*C*) Schematic of study workflow. (*D*) Genome-wide DNA accessibility and H3K27ac in adult human retina, macula, and retinal pigment epithelium (RPE) and choroid, by the assay for transposase accessible chromatin (ATAC), DNase-seq, and H3K27ac ChIP-seq. Each accessible genomic region (ATAC or DNase-seq) is represented as a single horizontal line centered on the peak summit, with a window of ±1 kb. ATAC or DNase-seq signal is plotted in gray. H3K27ac ChIP-seq signal is plotted in green. For each tissue, windows of DNA accessibility and H3K27ac are ordered on highest to lowest total H3K27ac signal within the 2-kb window. For retina analyses, ATAC $n = 8$ samples from unrelated adults, H3K27ac $n = 3$, and RNA-seq $n = 7$. For macula analyses, ATAC $n = 3$, H3K27ac $n = 3$, and RNA-seq $n = 3$. For RPE analyses, Dnase-seq $n = 2$, H3K27ac $n = 2$, and RNA-seq $n = 3$. (*E*) Expression of genes associated with DNA-accessible regions in adult human retina (seven individuals), macula (three individuals), and RPE/choroid (three individuals) as determined by sequencing total RNA from nuclei (Nuc-seq). Genes are ordered according to their corresponding accessible regions from *D*. (*F*) Representative gene loci showing custom UCSC browser tracks for ATAC-seq or DNase-seq, H3K27ac ChIP-seq, and total RNA Nuc-seq from adult human retina, macula, or RPE/choroid. (*G*) Enrichment of biological processes and phenotypes associated with candidate active CREs in each tissue according to analysis using the genome regions enrichment of annotations tool (GREAT) (24).

was expressed in both human RPE and retina (*SI Appendix*, Fig. S2*A*). One upstream enhancer (1) that is highly active in the retina (*SI Appendix*, Figs. S1 *E* and *F*, S2*A*, and S3 *D* and *E*) is accessible in RPE, but not enriched for H3K27ac. In contrast, two intronic enhancers (7′ and 8′) are accessible and enriched for H3K27ac in RPE/choroid, but not in retina or macula. Independent regulation of disease gene expression by tissue-specific CREs offers one possible explanation for variability seen in human disease phenotypes among individuals with commonly affected genes.

Whether tissue-specific CREs are primarily regulated at the level of accessibility or activation, the end result should be tissue-specific differences in gene expression. To confirm this, we first identified tissue-specific CREs by comparing levels of H3K27ac between each tissue (Fig. 2*E*). Next, we compared the expression of genes associated with these differentially active CREs (Fig. 2*F*). We found that, as a population, genes associated with tissue-specific CREs demonstrated a tissue-specific pattern of expression. These observations underscore the role of differentially active CREs in driving distinct programs of gene expression to impart unique biological functions to each tissue.

CRE Output Is Determined by Combinatorial TF Binding. Ultimately, TFs determine the accessibility, activity, and function of CREs. Furthermore, genetic variants within CREs that disrupt TF binding motifs may lead to misregulated gene expression and disease. It is therefore important to investigate the individual and combinatorial contributions of TFs to CRE function. We first sought to determine which TFs might bind to CREs in human retina, macula, and RPE/choroid by performing motif enrichment analysis using the HOMER software package (27). In the retina and macula, HOMER found significant enrichment of motifs matching the dimeric OTX2 and CRX (K50 homeodomain), MAF (basic region, leucine zipper), ROR (nuclear receptor), and MEF2 (MADS) TF family consensus binding motifs, whereas in RPE AP-1, MITF, TEAD, OTX2, and CRX motifs were enriched (Fig. 3*A*). Members of these TF families have evolutionarily conserved roles in retinal and/or RPE function and can lead to human visual disorders when mutated (14, 28). However, the combinatorial action of these TFs in the human retina is not well understood.

Motifs are only predictive of TF binding and do not unambiguously identify TFs, as multiple TFs may bind to similar motifs. Therefore, it is important to investigate directly how the binding and activity of TFs relate to human CRE function. We focused on the adult human retina and performed ChIP, followed by high-throughput sequencing (ChIP-seq) for five TFs

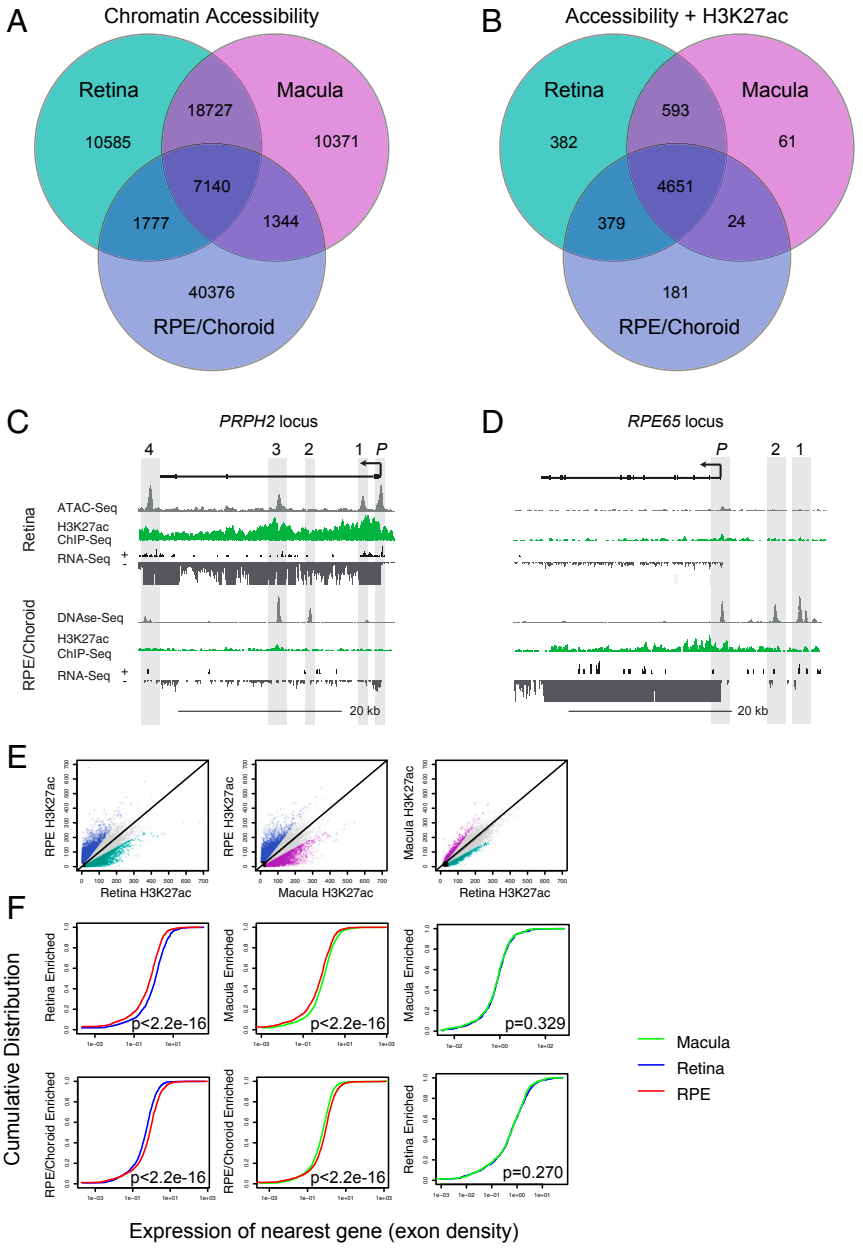

Fig. 2. Differential CRE accessibility drives unique patterns of gene expression in human retina, macula, and RPE. (*A*) Overlap of DNA-accessible genomic regions in adult human retina, macula, and RPE/choroid. (*B*) Overlap of H3K27ac-enriched genomic regions that share DNA accessibility in adult human retina, macula, and RPE/choroid. (*C* and *D*) Representative tissue-specific gene loci showing custom UCSC browser tracks for ATAC-seq or DNase-seq, H3K27ac ChIP-seq, and total RNA Nuc-seq from adult human retina and RPE/choroid. (*E*) Pairwise comparison of H3K27ac signal among DNA-accessible regions from adult human retina, macula, and RPE/choroid (blue, RPE enriched; green, retina enriched; purple, macula enriched). (*F*) Cumulative distribution of gene expression associated with tissue-specific DNA-accessible/H3K27ac regions.

whose motifs were highly enriched among retinal CREs: CRX, OTX2, NRL, RORB, and MEF2D. Antibodies were selected according to previous validation in the literature (8–11, 29). Three biological replicates were performed for each antibody unless two validated antibodies were available, in which case two biological replicates were performed per antibody. Anti-RORB antibody was only available in limited amounts, and therefore only the top 30% of total peaks were considered for further analysis that matched similar numbers of peaks found genome-wide with the other TFs. In each case, high confidence ChIP-seq peaks were identified across biological replicates using the ENCODE Consortium's irreproducible discovery rate (IDR) pipeline with a threshold of 1% (30). For each of the five assayed TFs, we found that TF protein binding was highly enriched at these active CREs in the adult human retina compared to shoulder regions and genome-wide chromatin input controls (Fig. 3*B*).

To validate the specificity of these antibodies, we performed reciprocal motif analysis for all regions bound by each TF according to ChIP-seq. In each case, we found that the specific motif corresponding to the antibody-targeted TF was significantly enriched (*SI Appendix*, Fig. S3*A*). Additionally, we found that the K50 homeodomain TF motif (e.g., CRX, OTX2) was also highly enriched for all TF-bound regions, suggesting that CRX and/or OTX2 cobind with these other TFs at human retinal CREs. Alternatively, this motif may simply be indiscriminately enriched at sites of binding for any TF in the retina. To distinguish between these possibilities, we performed ChIP-seq for additional factors CTCF and CREB that are not known to directly regulate photoreceptor cell gene expression. Regions bound by these TFs were highly enriched for their own motifs but did not demonstrate enrichment for the K50 homeodomain consensus motifs, suggesting that CRX and OTX2 are acting on a specific subset of retinal CREs (*SI Appendix*, Fig. S3*A*). These findings are further supported by previous studies of CRX and CTCF binding in the mouse retina (31, 32). Taken together, these results support the specificity of the antibodies used for these experiments and underscore the combinatorial function of CRX and OTX2 with other TFs in establishing and maintaining photoreceptor enhancers (8, 33).

To examine the relationship of individual TF binding and CRE activity, we ranked human retinal CREs from most active to least active according to H3K27ac enrichment (Fig. 3*B*). We found that the binding of some TFs (OTX2, NRL, MEF2D) was positively correlated with H3K27ac levels, suggesting that these TFs or combinations of these TFs may be activating. Oppositely, binding of CRX was highest at the bottom quintile of H3K27ac levels. This suggested that CRX can mediate CRE repression in the mature human retina or is at least bound to inactive sites in the genome (Fig. 3*B* and *SI Appendix*, Fig. S3 *B–D*). To further examine these sites, we performed ChIP-seq for an additional histone mark H3K4me2, which is enriched at CREs that are either active (H3K27ac-enriched) or potentially poised (H3K27ac-low) (Fig. 3*B*). Sites at which CRX is bound, but H3K27ac is low are still positive for H3K4me2, suggesting that these sites are silent or poised CREs. Furthermore, a repressive role for CRX at some CREs is consistent with a recent study in the mouse retina using massively parallel reporter assays (33). In future studies, additional support for this hypothesis could be gained by comparing CRX binding to marks of repressive chromatin such as H3K27me3. At present, however, this study, as well as others (8), demonstrates that when CRX binds together with other TFs, the effect can be activating. These observations underscore the significance of combinations of TFs in regulating CREs to promote or repress gene expression.

To assess the significance of specific combinations of TF binding to different aspects of CRE regulation, we determined levels of DNA accessibility, H3K27ac enrichment, and nearby gene expression for each permutation of TF binding for five TFs (*SI Appendix*, Fig. S3 *B–D*). Using the Brown–Forsythe ANOVA test with a Dunnett correction for multiple comparisons, after removing outliers, we identified significant differences among differentially bound CREs for each feature (ATAC: $F = 484.9$ [df = 29,603], $P < 0.0001$; H3K27ac: $F = 157.5$ [df = 39,921], $P < 0.0001$; RNA: $F = 83.98$ [df = 32,454], $P < 0.0001$). Among significant pairwise differences, we found that, as a population, sites where CRX binds without the other assayed TFs have significantly lower levels of DNA accessibility, lower levels of H3K27ac, and lower levels of associated gene expression. These sites, however, are enriched for H3K4me2, the epigenetic mark of enhancers and promoters (Fig. 3*B*). This suggests that such sites may be maintained in a poised manner, possibly as a vestige of development or as CREs primed to mediate a stimulus-dependent transcriptional response. We did not identify consistent trends across other specific combinations of TF binding that correlated with levels of accessibility, H3K27ac, or nearby gene expression. Taken together, this would suggest that regulation of human retinal CREs is heterogeneous and flexible as opposed to fixed according to an invariant order of TF binding. Among models of CRE regulation that have been proposed, these data are more consistent with a flexible "billboard" model of TFs that regulates gene expression rather than an enhanceosome with a fixed spacing and configuration of TF binding and regulation (34).

To further investigate the significance of TF cobinding and to examine the DNA sequence requirements of human CREs, we returned to the *ABCA4* locus. Determining sequence requirements for CREs is key to interpreting the effect of genetic variants within CREs that may cause disease. Multiple TFs bound to a single CRE may additively or synergistically regulate target gene expression. Alternatively, these TFs may act in a redundant manner to ensure robustness against genetic or cellular perturbations. To distinguish among these possibilities, we mutated individual TF binding motifs at this single, highly active enhancer that binds all five of the TFs that we have examined (Fig. 3*D*). Instead of ensuring robustness, we found that disruption of individual motifs had a profound effect on CRE activity (Fig. 3*E*). Even motifs that occurred multiple times within the same CRE like the shared K50 homeodomain motif (CRX/OTX2) were not redundant. The only motif that did not appear to be required for full CRE function was the nuclear receptor (likely bound by RORB) motif. These observations are notable as they show that each TF binding event at this highly active enhancer is contributing in a nonredundant manner to enhancer activity, and that CRX appears to function as a transcriptional activator in this context. Taken together, we find that the combination of TFs that bind to individual enhancers can significantly affect CRE activity. This activity may be regulated at the level of DNA accessibility, or CRE activation, and can interact with additional CREs to affect target gene expression. These observations are informative for interpreting the impact of genetic variation within human CREs.

Deconvolution of Cell Type-Specific Gene Regulation. The signal arising from epigenomic assays performed on the intact retina is likely a function of the abundance of each cell type within the tissue and the strength of the signal within that cell type. Nonetheless, cell type-specific CREs may still be identified by focusing attention on CREs associated with cell type-specific gene expression (35). The abundance of different cell types in the human retina and the patterns of cell type-specific gene expression, however, have not been extensively characterized. Single-cell RNA-seq is a powerful tool to deconvolve cell types in complex tissues including the retina (36–39). To identify cell type-specific CREs, we first sought to determine the abundance of different cell types in the human retina and to identify cell type-specific gene expression in human retinal cells. Then, with knowledge of the genes that are selectively expressed in a given cell type, it should be possible to assign the nearby CREs to that cell type.

We optimized the protocol of Klein et al. (37) to profile individual cell nuclei (single-cell Nuc-seq) from adult human retinas from three unrelated individuals (Fig. 4*A*). By profiling RNA from single nuclei, we avoid transcriptional contamination caused by cell-to-cell contacts and obtain a more accurate snapshot of active transcription. This also obviates the need for intact, living cells because intact nuclei can be efficiently isolated from flash-frozen human tissue, making it possible to use stored tissue from biobanks. Our single-cell sequencing data were analyzed using Seurat, version 3.0 (40, 41). After cell-free droplets, doublets, and cells with high mitochondrial reads were removed, 4,763 cells remained. We then identified clusters of cells that, based on expression of known marker genes, corresponded to all major cell classes of the mature retina including rod and cone photoreceptors, horizontal, bipolar, Müller glial, amacrine, and retinal ganglion cells (Fig. 4*A*) (42–47). These classes were identified retrospectively based on patterns of known marker gene expression (*SI Appendix*, Fig. S4*A*). Rod photoreceptor cells were the most abundant class of cells profiled, suggesting that the majority of the ATAC-seq and ChIP-seq signal originates from epigenomic features in this cell type (Fig. 4*B*). However, the abundance of other cell classes in the human retina was higher than that of nonrod cells observed from single-cell analysis of the mouse retina (36). Notably, cone photoreceptor and horizontal cells made up a larger percentage of all cells in the human retina compared to mouse, consistent with classical neuroanatomical studies of rodents and primate and rodent retinas (48).

To determine whether candidate cell class-specific CREs could be identified, we verified the expression of cell class-specific marker genes in our single–Nuc-seq dataset (*SI Appendix*, Fig. S4*A*). We then examined these individual loci in our dataset to assess DNA accessibility, histone modifications, and TF binding (*SI Appendix*, Fig. S4*B*). We found that the following cell class markers maintained their specificity of enrichment in the human retina and had observable signatures of CREs at their gene loci: *RP1* and *NR2E3* (rod photoreceptors), *ARR3* and *PDE6H* (cones), *ONECUT2* (horizontal cells), *VSX2*, *PRKCA*, and *GRM6* (bipolar cells), *CLU* and *APOE* (Müller glial cells), *GAD1* (amacrine cells), *SCL17A6*/*NEFL* (Ganglion cells), and *GFAP* (astrocytes) (Fig. 4 and *SI Appendix*, Fig. S4). These genes represent markers of broad classes of cells in the retina. To determine whether CREs associated with lowly abundant cell types were identifiable in our dataset, we analyzed gene loci of cone photoreceptor cell type markers *OPN1LW* and *OPN1SW*. Each of these gene loci demonstrated nearby accessible regions. In the case

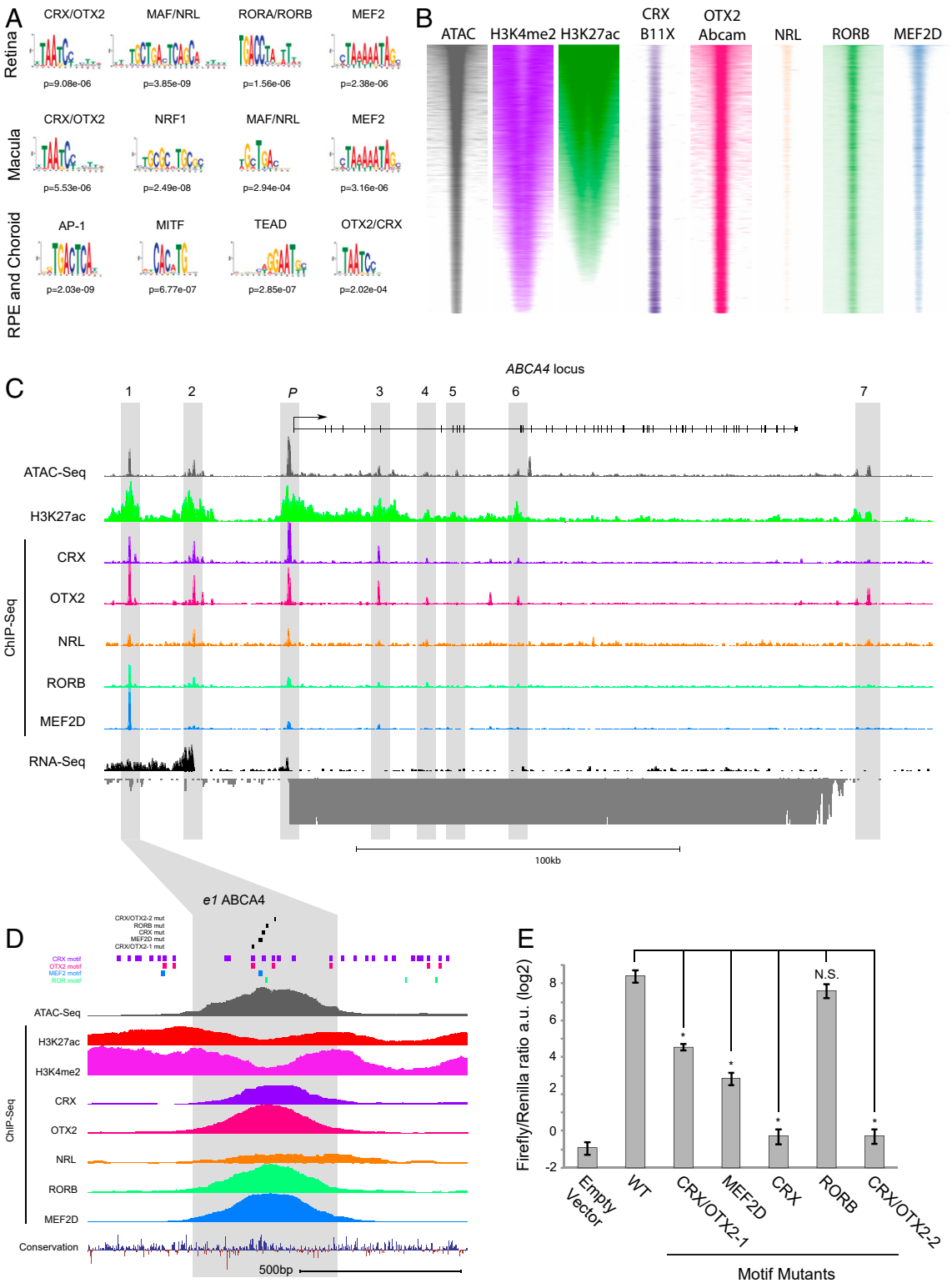

Fig. 3. Coregulation of human retinal enhancers by combinatorial binding of TFs. (*A*) Position weight matrices (PWMs) of TF binding motifs enriched within accessible/H3K27ac+ genomic regions from adult human retina, macula, and RPE/choroid. (*B*) Genome-wide distribution of DNA accessibility, H3K4me2, H3K27ac, and TF binding in adult human retina by the assay for transposase accessible chromatin (ATAC) ($n = 7$) and H3K4me2 ($n = 3$), H3K27ac ($n = 3$), or TF ChIP-seq (see *Materials and Methods* for *n*). Each genomic region is represented as a single horizontal line centered on the peak summit, with a window of ±1 kb. Genomic regions are ordered on highest to lowest total H3K27ac signal. (*C*) ABCA4 gene locus showing custom UCSC browser tracks for ATAC-seq, H3K27ac, and TF ChIP-seq and total RNA Nuc-seq from adult human retina. Individual candidate CREs (*e1-e7*, *promoter*) are highlighted in gray. (*D*) A candidate enhancer (*e1*) upstream of the ABCA4 gene showing custom UCSC browser tracks for TF binding motif mutations (black), TF binding motifs (purple, CRX; red, OTX2; blue, MEF2; green, ROR), ATAC-seq, H3K27ac, H3K4me2, and TF ChIP-seq and vertebrate conservation from adult human retina. Area highlighted in gray corresponds to sequence assayed in *E*. (*E*) Luciferase reporter assay comparing activity of the consensus human enhancer sequence highlighted in *D* to induced mutations of individual TF binding motifs (*$P < 0.05$).

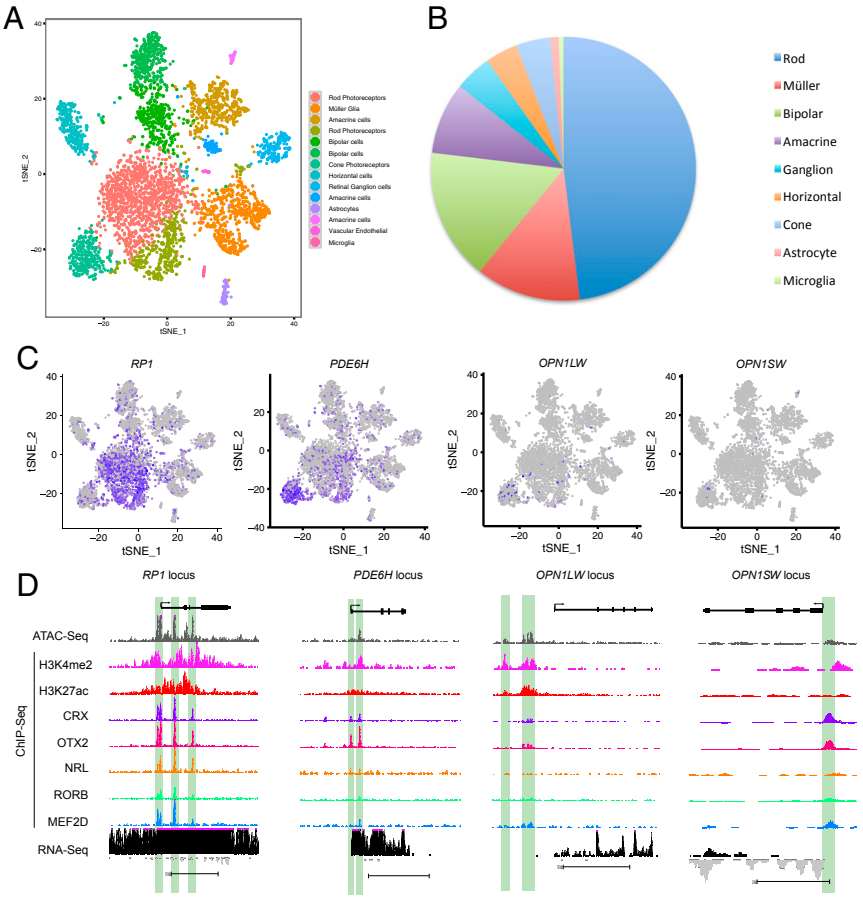

Fig. 4. Deconvolution of cell type-specific regulatory elements. (*A*) t-Stochastic neighbor embedding (tSNE) plot of global gene expression of 4,763 single nuclei from adult human retinal cell nuclei; nuclei are colored according to 14 unsupervised clusters. (*B*) Proportion of cell classes in the human retina identified according to known marker gene expression in single retinal nuclei (n = 3 biological replicates) (*SI Appendix*, Fig. S4*A*). (*C*) Marker gene expression of individual photoreceptor cell types (*RP1*, rod; *PDE6H*, pan-cone; *OPN1LW*, long-wavelength [red] cones; *OPN1SW*, short-wavelength [blue] cones). (*D*) Epigenetic features at photoreceptor cell type-specific marker gene loci shown by custom UCSC browser tracks for ATAC-seq, H3K4me2, H3K27ac, and TF ChIP-seq and total RNA Nuc-seq from adult human retina.

of *OPN1SW*, the promoter region was accessible with binding of CRX, OTX2, and MEF2D. At the *OPN1LW/MW* region, the well-studied upstream locus control region (LCR) (49) was clearly accessible and binding of CRX, OTX2, and MEF2D was observed (Fig. 4*D*). Together, these findings suggest that candidate cell class- and cell type-specific regulatory elements can be identified in these integrative datasets, although they are generated using heterogeneous retinal tissue, even from cell types that represent a small fraction of the total tissue. Newly available single-cell assays of DNA accessibility and histone modifications should ultimately be used to directly test the proposed specificity of these CREs.

To examine the profile of cell class-specific genes associated CREs as a population, we identified the top 100 most-enriched genes expressed in each cell class whose promoter regions demonstrated accessibility by ATAC-seq (*SI Appendix*, Fig. S4*C*). We then quantified the ATAC-seq and H3K4me2 and H3K27ac ChIP-seq signal at the promoters of these genes (*SI Appendix*, Fig. S4*C*). For each cell class-enriched cohort, we observed enriched H3K27ac and H3K4me2. Together, these data suggested that epigenetic features of cell class-specific promoters can be observed in datasets collected from intact tissue.

Conservation and Evolution of Cell Type-Specific Regulatory Elements. CREs are important substrates of evolutionary change (3). To assess whether conserved retinal architecture is reflected in the genome in the form of conserved CREs, we examined the conservation between human and mouse CREs at critical visual genes. We first compared the number and topography of candidate enhancers at the mouse and human Rhodopsin locus. Rhodopsin is a G-protein–coupled receptor, essential for rod cell phototransduction in all mammals. It is also among the most commonly mutated genes in human inherited retinal disorders (12). We observed remarkable conservation of regulatory element topography between mice and humans despite extensive evolutionary divergence between the two species (Fig. 5*A*). This topographic conservation is evident from the number and relative location of individual CREs at the *RHO* locus. Additionally, the pattern of TF binding to each CRE appears also to be conserved with the exception of MEF2D, which appears to be present at CREs at the human, but not mouse *RHO* locus. Despite this topographic conservation, the conservation of DNA sequence is limited to discrete regions within each CRE (*SI Appendix*, Fig. S5*A*). This suggests that regulatory conservation can be maintained in the absence of extensive DNA sequence conservation and that more specific conservation at the level of individual TF binding sites is likely what maintains functional conservation.

To assess conservation of regulatory elements at a gene expressed in a less abundant cell type, we analyzed the *CABP5* locus, a calcium binding protein that is an established marker of bipolar cells in the mouse retina (50). *CABP5* was originally identified as sharing homology to calmodulin and is closely related to *CABP4*, which is associated with inherited retinal disorders (51, 52). A proximal promoter and a downstream enhancer are evident in both mouse and human retinas (Fig. 5*B*). The topography of these elements with respect to the *CABP5* coding sequence is conserved; however, the binding of TFs is not. Only OTX2 was bound to the mouse promoter and enhancer, while the human promoter and enhancer have acquired binding of CRX, OTX2, NRL, MEF2D, and RORB. Among these factors, CRX and NRL are enriched in rod photoreceptors, suggesting that this gene may be differentially regulated in the human retina. To determine whether the regulation of *CABP5* has been altered between the mouse and human retina, we analyzed our human single-cell transcriptome data and compared these to single-cell data acquired from the mouse retina (36). In mouse, *Cabp5* expression is largely restricted to bipolar cells, whereas it is

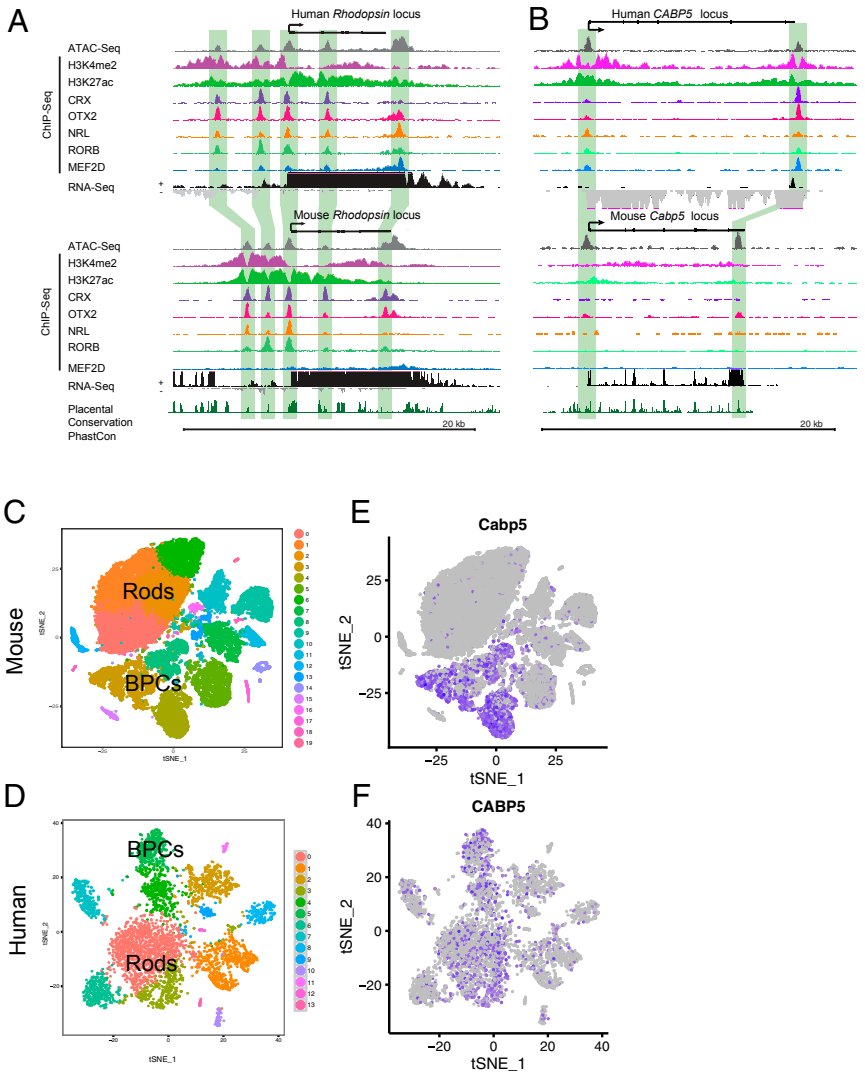

Fig. 5. Conservation and divergence of CRE function in the human and mouse retina. (*A*) *Rhodopsin* gene locus showing custom UCSC browser tracks for ATAC-seq, H3K4me2, H3K27ac, and TF ChIP-seq and total RNA Nuc-seq from adult human or mouse retinas. Individual candidate CREs are highlighted in green. (*B*) *CABP5*/*Cabp5* gene locus showing custom UCSC browser tracks for ATAC-seq, H3K4me2, H3K27ac, and TF ChIP-seq and total RNA Nuc-seq from adult human or mouse retinas. Individual candidate CREs are highlighted in green. (*C* and *D*) t-Stochastic neighbor embedding (tSNE) plot of 31,528 single cells from the mouse retina from ref. 36 (*C*) or 4,763 single-cell nuclei from adult human retinas (*D*) according to global gene expression; nuclei are colored by unsupervised clusters (BPCs, clusters composed of bipolar cells; rods, clusters composed of rod photoreceptor cells). (*E* and *F*) Expression of *Cabp5*/*CABP5* among single retinal cells/nuclei in the mouse and human retina, respectively.

expressed in bipolar cells and rod photoreceptors in the human retina (Fig. 5 *C*–*F*). Of note, a 54-bp low-complexity (CT) insert in the downstream mouse enhancer may prevent the recruitment of other TFs (*SI Appendix*, Fig. S5*B*). Together, these observations show that interspecies differences in gene expression can occur through sequence divergence in promoters and enhancers to affect not just the activity of a gene, but also its cell type specificity.

Developmentally Dynamic CREs Underlie Distinct Biological Functions, Pathologies, and Regulatory Mechanisms. CREs are highly dynamic during tissue development and regulate discrete processes including proliferation, specification, and differentiation (35, 53). To investigate the genetic regulation of these processes in the human retina, we sought to identify developmentally dynamic CREs. Furthermore, visual disorders can arise from gene misregulation during retinal development as well as in the adult retina. It is therefore important to consider developmentally dynamic CREs in conjunction with adult CREs when searching for disease-causing variants.

At 10 to 12 wk of human fetal development (70 to 84 d), the earliest born cell types, the retinal ganglion cells and cone photoreceptors, are beginning to differentiate. By 14 to 17 wk, most neurons of the retina have left the cell cycle and the latest-born cell types including rod photoreceptors are differentiating (98 to 119 d) (54, 55). We compared chromatin accessibility in the adult human retina to that of the developing retina at these two critical stages (20, 53).

Clustering of chromatin accessibility data from these distinct time points revealed three specific patterns that dominated the landscape of developmentally dynamic DNA accessibility (Fig. 6*A*). The majority of sites observed at any time point was shared between the two developmental time windows but was closed in the adult retina (21,828/44,510: 49%). For example, enhancers upstream of the *ATOH7* gene locus, which is required for the differentiation of early-born retinal ganglion cells and has recently been shown to regulate aspects of rod vs. cone cell fate in the human retina, exhibited this pattern of chromatin accessibility (Fig. 6*B*) (16, 56). There was also a substantial fraction of CREs that were constitutively open across all time points (16,302/44,510: 36.63%) and CREs that were only accessible in the adult differentiated retina (5,518/44,510: 12.40%), such as those found at the *RHO* locus (Fig. 6*C*). Very few regions were specific to either developmental window (10 to 12 wk, 378 [0.85%], or 14 to 17 wk, 346 [0.78%]) or shared between the late developmental window and adult retina (110 [0.25%]). As expected, almost no sites were accessible in a discontinuous pattern, i.e., present at 10 to 12 wk and adult, but absent from 14 to 17 wk (28 [0.06%]). These patterns suggest highly transient transcriptional programs that are active in human retinal development and demonstrate that CREs continue to be licensed and decommissioned as cells mature.

Identification of dynamic chromatin accessibility is necessary to completely define the genomic search space for CRE variants that cause disorders of human visual development. To explore

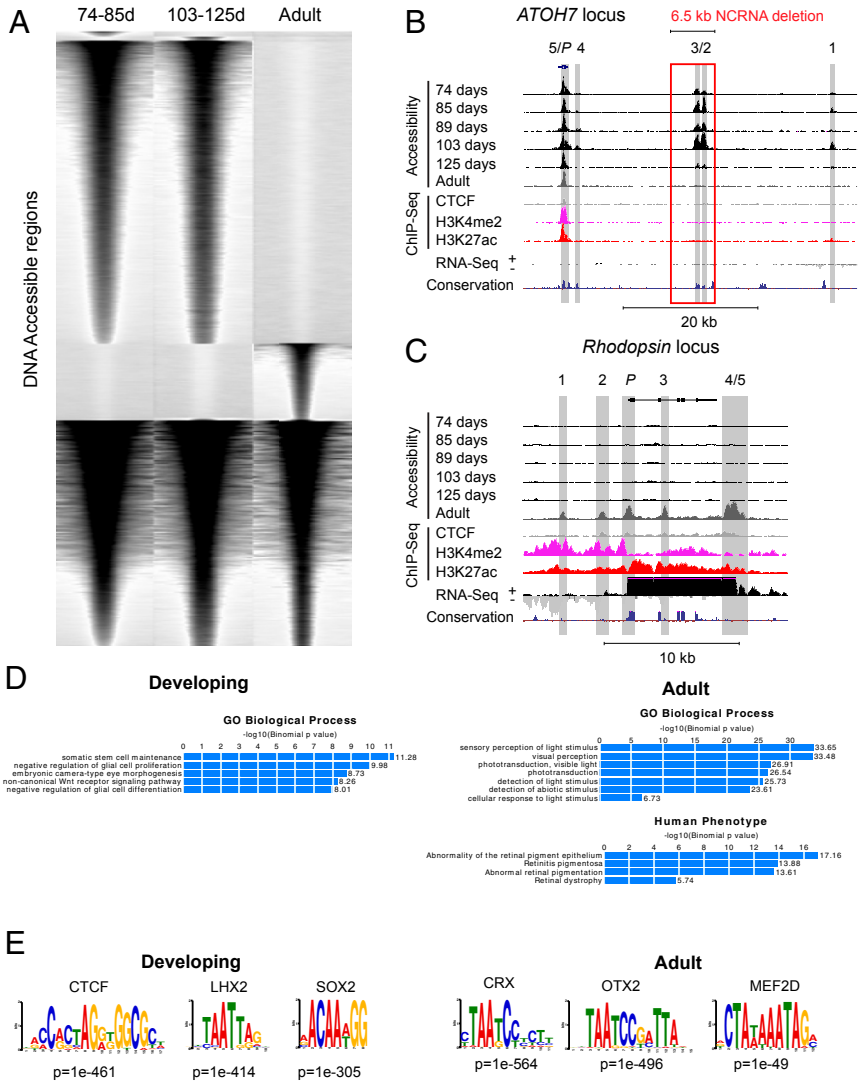

Fig. 6. Developmentally dynamic DNA accessibility underlies distinct biological functions, pathologies, and regulatory mechanisms. (*A*) Dynamic DNA accessibility during human retinal development. Each accessible genomic region (ATAC or DNase-seq) is represented as a single horizontal line centered on the peak summit, with a window of ±1 kb. ATAC (adult) or DNase-seq (developing) signal is plotted in black. For each developmental stage, windows of DNA accessibility are ordered on highest to lowest total accessibility signal. (*B* and *C*) Developmentally dynamic DNA accessibility at the *ATOH7* (*B*) or *Rhodopsin* (*C*) gene loci displayed as custom UCSC browser tracks for DNase-seq (developing) or ATAC-seq (adult) from a time course of human retinal development and CTCF, H3K4me2, and H3K27ac ChIP-seq and total RNA Nuc-seq from adult human retina (red box, 6.5 kb deleted ATOH7 enhancer region resulting in inherited human nonsyndromic congenital retinal nonattachment; NCRNA; ref. 16). (*D*) Enrichment of biological processes and phenotypes associated with developmentally dynamic candidate CREs in the human retina according to analysis with the genome regions enrichment of annotations tool (GREAT). (*E*) TF binding motif enrichment at regions that are accessible in developing or adult human retina.

this further, we examined developmental chromatin accessibility at two CREs associated with human disease-causing enhancer mutations. Individuals who are homozygous for a deletion of a distal enhancer of *ATOH7* are blind from birth and demonstrate congenital retinal nonattachment (16). Importantly, we found that there are actually two regions of DNA accessibility in the developing human retina encompassed within this deleted region, suggesting that two, not just one, enhancer elements are lost in this disorder (Fig. 6*B*). Additionally, an individual with a de novo monoallelic mutation at a PAX6 autoregulatory site demonstrates the congenital eye malformation aniridia and absence of a fovea (17). We found that this point mutation also falls within a region of DNA accessibility present during development (*SI Appendix*, Fig. S6*A*). Neither of these enhancers demonstrates accessibility in the adult retina, suggesting that a distinct cohort of genomic regions should be used to search for genetic defects causing developmental disorders rather than disorders that manifest later in life. In contrast, a third disease-associated variant, single-nucleotide polymorphism (SNP) rs17421627, has been shown to cause developmental retinal vascular abnormalities and is strongly associated with macular telangiectasia type 2 (57, 58). This point mutation falls within an enhancer that is accessible both developmentally and in the adult retina, suggesting how it can contribute to both developmental and later-onset retinopathies (*SI Appendix*, Fig. S6*B*).

The distinct patterns of chromatin accessibility in the developing vs. the adult retina further suggest that these elements serve different biological functions and may impinge upon distinct pathologies. We found that regions of accessible chromatin in the developing retina were significantly associated with the biological processes of stem cell maintenance, embryonic camera-type eye morphogenesis, and negative regulation of glial development. Adult-specific accessibility, however, was significantly associated with light detection and phototransduction. Apart from DNA accessibility at the *ATOH7* and *PAX6* enhancers, no human disease phenotype terms were associated with regions that are open exclusively in the developing retina. In contrast, adult-specific accessible regions were significantly associated with abnormality of the RPE, retinitis pigmentosa, abnormal retinal pigmentation, and retinal dystrophy (Fig. 6*D*).

Distinct patterns of DNA accessibility across retinal development are likely driven by distinct cohorts of TFs. To identify potential regulators of these distinct elements, we looked for known TF DNA binding motifs that were enriched within developmental- or adult-specific accessible regions or within regions that are constitutively accessible. Among the top motifs enriched for developmental-specific accessibility are those that are assigned to CTCF, LHX, and SOX family TF (Fig. 6*E*). LHX and SOX TF families have well-documented roles in the maintenance of progenitor cells during retinal development (59). CTCF is best known for its association with transcriptional

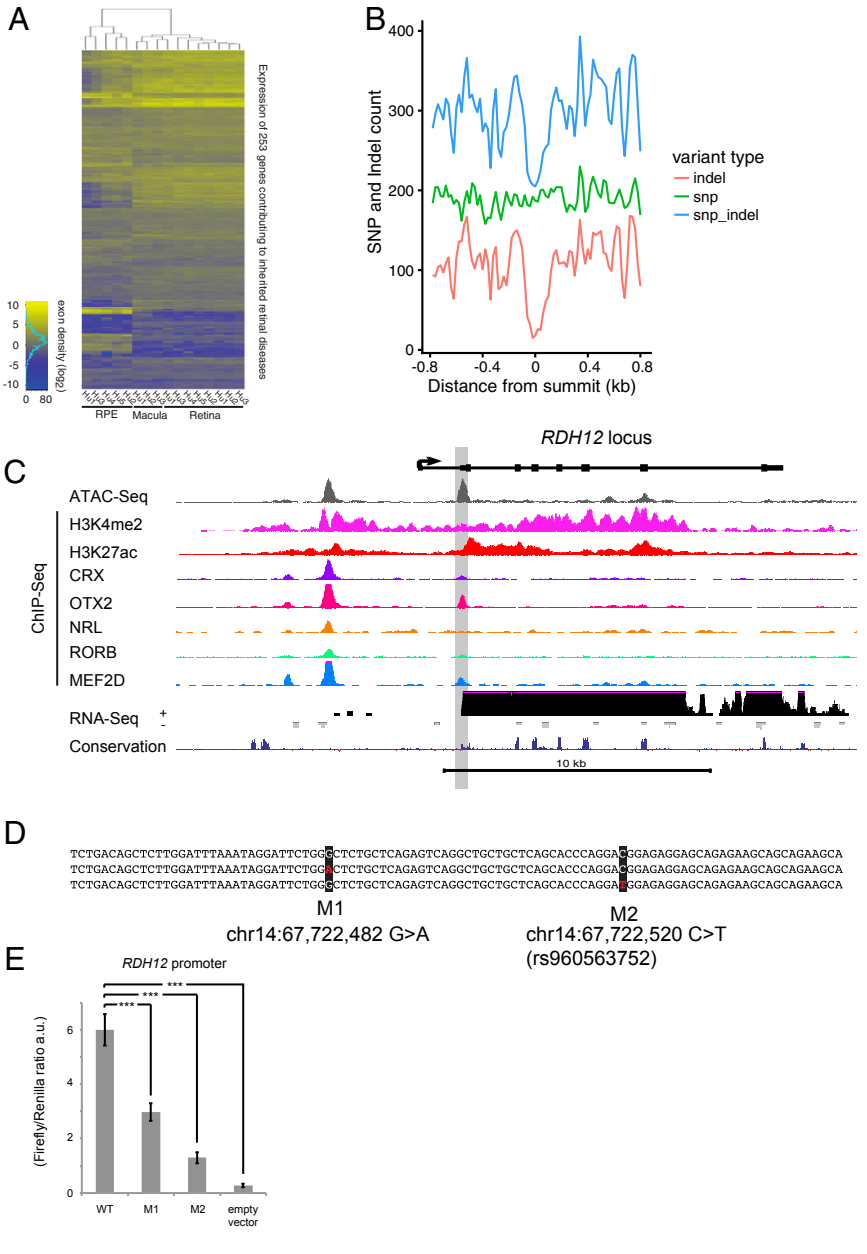

Fig. 7. Disease gene-associated CREs and noncoding variation. (*A*) Heatmap of human retinal disease gene expression by tissue type. (*B*) SNP and indel frequency at disease gene-associated CREs (gnomAD, MAF 0.01) (8,685 indels, 15,306 SNPs). (*C*) The *RDH12* gene locus showing custom UCSC browser tracks for ATAC-seq, H3K4me2, H3K27ac, and TF ChIP-seq and total RNA Nuc-seq from the adult human retina. Presumptive promoter region highlighted in gray. (*D*) Consensus DNA sequence centered around promoter ATAC-seq peak summit. Single-nucleotide variants (SNVs) (red with black highlight) found in unrelated individuals with Leber congenital amaurosis/retinitis pigmentosa. (*E*) Luciferase reporter assay comparing the relative activities of RDH12 consensus and variant promoter constructs (***$P < 0.0005$).

insulators and is also the topmost enriched motif within the constitutively accessible regions. In the adult-specific accessible regions, the top-most enriched motifs are assigned to the TFs CRX, OTX2, MEF2D, and ROR, which are all known to be essential for proper photoreceptor cell differentiation and function (Fig. 6*E*). Consistent with these analyses, we established that the patterns of TF binding were highly concordant with the motif enrichment and specificity of these distinct regions (*SI Appendix*, Fig. S6*C*). CRX, OTX2, MEF2D, and RORB were more frequently bound at adult-specific regions compared to constitutively accessible regions. CTCF, in contrast, was more robustly bound to constitutively accessible regions compared to regions specifically accessible in the adult retina.

Taken together, these findings show that unique networks of TFs function at dynamic CREs to regulate stage-specific gene biological functions, whereas the binding of CTCF at putative insulator elements appears to be largely static and is specified early in human retinal development. Furthermore, these data demonstrate that identified CREs in the human retina overlap with known disease-causing regulatory element mutations. This observation serves as an important proof of concept that mapping CREs can help define the genomic search-space for novel variants that may influence human vision.

Identification and Characterization of Functional Variants within Human *cis*-Regulatory Regions. To prioritize the search for novel noncoding variants that affect human vision, we first identified CREs associated with known retinal disease genes (12) (Fig. 7*A*). Of 253 disease genes, we identified 763 enhancers and promoters (Dataset S3). Genetic variants within these regulatory elements could contribute to disease phenotypes by misregulating the expression of these disease-associated genes.

If genetic variants within these CREs are deleterious, then such variants should be rare in the general human population. In support of this concept, a recent study demonstrated a depletion of variants at CRX-bound sites between genetically divergent mouse strains (60). To assess this in humans, we determined the frequency of SNPs and indels (insertions and deletions) within human retinal-disease gene CREs from ∼15,500 unrelated individuals without visual disorders (61). We found that indels were strongly depleted in

the 200 bp centered on the peaks of chromatin accessibility compared to flanking regions (Fig. 7*B*). In contrast, no significant depletion of SNPs was observed, consistent with a higher tolerance for single base changes at CREs. Even so, several reports demonstrate that single base changes within CREs can have profound effects on enhancer function and disease phenotypes (17, 57, 62).

To identify novel candidate functional variants in human retinal CREs, we queried the sequence within these regions from individuals with defined retinal pathologies. These individuals were selected as potentially carrying noncoding regulatory variants because they were found to be heterozygous for a recessive coding pathogenic variant or loss-of-function allele in a known disease gene that matched the disease phenotype. Our hypothesis was that genetic variation within CREs associated with these genes might contribute to the disease phenotype by further lowering the effective dosage of the gene product. For several different unrelated individuals, we identified noncoding variants in the CREs of known retinal disease genes including *RDH12*, *ABCA4*, and *MYO7A*.

Two such variants were identified in an active CRE at the *RDH12* locus in individuals with inherited retinal disease. Biallelic mutations within the *RDH12* coding sequence are known to cause recessive Leber congenital amaurosis with severe childhood retinal dystrophy (63). We found two active CREs at the *RDH12* locus (Fig. 7*C*). The structure of this gene has been predicted to span eight exons with the first exon being noncoding (RefSeq NM_152443). We found no evidence of CRE accessibility or significant DNA sequence conservation at the predicted promoter. Instead we found accessibility at the start of the predicted second exon near the start of the *RDH12* ORF. Together with our total RNA-seq reads, these data suggested an alternative promoter for *RDH12* in the adult human retina. At this candidate promoter, we observed local enrichment for H3K27ac and binding of CRX, OTX2, RORB, and MEF2D, albeit at moderate levels compared to a nearby upstream enhancer (Fig. 7*C*).

At this prospective promoter, we found two separate variants in two unrelated individuals, *RDH12 c.-162G>A* (hg38, chr14:67722482G>A) and *RDH12 c.-123C>T* (hg38, chr14:67722520C>T) (reported in ref. 64) (Fig. 7*D*). Neither of the variants was previously characterized as deleterious, however *c.-123C* is conserved across simians and *c.-162A* is conserved down to monotremes. We tested the effect of these variants on the transcriptional activity of this CRE. The consensus CRE demonstrates substantial promoter activity in a plasmid-based reporter assay. However, when each SNP is introduced individually, the activity of this region decreases by ~50% and 80%, respectively ($P < 0.001$, $n = 12$, t test, two-tailed, homoscedastic) (Fig. 7*E*).

We tested additional identified variants in the *ABCA4* promoter and an intronic *ABCA4* enhancer and at an intronic *MYO7A* enhancer. Several SNPs assayed at the *ABCA4* promoter had no discernable effect on reporter activity. Surprisingly, however, two variants (rs11802887 and rs11806223) that were found together in siblings affected by Stargardt disease demonstrated increased reporter activity individually, with one increase reaching significance (rs11802887) ($P < 0.05$, $n = 4$, t test, two-tailed, homoscedastic) (*SI Appendix*, Fig. S7*A*). In contrast, an *ABCA4* enhancer variant (rs752024867; hg38 chr1:94,079,815T>G) found in an unrelated individual with Stargardt disease showed a significant decrease in reporter activity ($P < 0.05$, $n = 4$, t test, two-tailed, homoscedastic) (*SI Appendix*, Fig. S7*B*). Finally, a distinct SNP found in a *MYO7A* enhancer (chr11:77,172,048C>T) also demonstrated a significant decrease in enhancer activity ($P < 0.01$, $n = 4$, t test, two-tailed, homoscedastic) (*SI Appendix*, Fig. S7*C*). A common theme among all of these variants is that they do not fall within readily recognizable TF binding motifs; however, they do demonstrate conservation across clades. This suggests that despite reduced conservation in CREs compared to coding regions, conservation may be predictive of vulnerable DNA sequences within CREs that future studies could attribute directly to disease phenotypes (3, 57).

Discussion

The majority of human genetic variation falls within noncoding regions of the genome; however, these regions remain largely uncharacterized and are not routinely screened for disease-causing mutations. Genome-wide assays to identify CREs based on epigenetic signatures have greatly improved our understanding of functional elements within the noncoding genome. In this study, we identify CREs that are active in the adult human retina, macula, and RPE/choroid, three tissue regions that are required for vision and that are frequently affected in visual disorders. We find that many of these CREs demonstrate tissue-specific DNA accessibility and each tissue is enriched for a unique cohort of TF binding motifs. Within the retina, we mapped the binding of five TFs known to be required for photoreceptor function and recognized patterns of TF binding associated with CRE activity. We identified putative cell type-specific CREs by generating single-cell RNA-seq datasets to compare to our integrative epigenomic data. Furthermore, our studies characterize CREs that harbor noncoding variants that are associated with inherited visual disorders.

Our study focuses on adult human eye tissues and yields insights into genetic regulation, complementing previous studies in the mouse retina and the human retina (8–11, 35, 53, 65–67). By comparing human and mouse data together, our studies shed light on the evolution of noncoding regulatory elements in the visual system, a remarkably well-conserved part of the CNS. By comparing CRE landscapes derived from developmental and adult datasets, it is possible to identify distinct CREs that drive gene expression at each stage in retinal development and to map the binding of TFs that constitute the core transcriptional regulatory circuitry of retinal differentiation. Perhaps most importantly by comparing between developing and adult CREs in the human retina, it is possible to identify and test novel and known genetic variants that are associated with developmental and later-onset human visual diseases. Taken together, our data demonstrate the utility of this integrative genomic approach for the characterization of three complex human tissue regions that facilitate vision. This work fills important gaps in identifying and characterizing the function of noncoding regulatory elements and in providing a biologically relevant framework by which genetic variants can be evaluated and understood.

Materials and Methods

Human and mouse tissue was procured according to protocols approved by the Harvard Medical Area Institutional Review Board (IRB15-1457) and the Harvard Medical Area (04572) and the Seattle Children's Research Institute (00050) Institutional Animal Care and Use Committees. High-throughput epigenomic assays and luciferase assays were performed according to published protocols. More detail regarding specific experimental procedures and analyses are provided in *SI Appendix*.

Data Availability. All sequencing data generated for this study have been deposited in the GEO database (accession no. GSE137311).

ACKNOWLEDGMENTS. We thank Janine Zieg for assistance with figures, and members of the T.J.C. and M.E.G. laboratories for helpful discussions and suggestions. Liesl Strand and James Gillespie provided technical assistance with the luciferase assay for the 5q14.3 enhancer. Sinisa Hrvatin provided in-depth assistance with the single-nuc-RNA-sequencing. We also thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. This work was funded in part through the National Institute of Neurological Disorders and Stroke Javits Award to M.E.G. (R37 NS028829), laboratory startup funds to T.J.C., the Ghent University Special Research Fund (BOF15/GOA/011) to E.D.B., the Hercules Foundation AUGE/13/023 to E.D.B., the Research Foundation Flanders (Grants G0C6715N and G0A9718N) to E.D.B., the Foundation JED to E.D.B., and Funds for Research in Ophthalmology to M.B. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the funding sources mentioned.

1. A. S. Nord *et al*., Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
2. M. T. Maurano *et al*., Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
3. D. Villar *et al*., Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
4. C. T. Gordon, S. Lyonnet, Enhancer mutations and phenotype modularity. *Nat. Genet.* **46**, 3–4 (2014).
5. M. Spielmann, S. Mundlos, Looking beyond the genes: The role of non-coding variants in human disease. *Hum. Mol. Genet.* **25**, R157–R165 (2016).
6. F. Zhang, J. R. Lupski, Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
7. E. Birney *et al*.; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
8. M. M. Andzelm *et al*., MEF2D drives photoreceptor development through a genome-wide competition for tissue-specific enhancers. *Neuron* **86**, 247–263 (2015).
9. H. Hao *et al*., Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL targetome analysis. *PLoS Genet.* **8**, e1002649 (2012).
10. A. Samuel, M. Housset, B. Fant, T. Lamonerie, Otx2 ChIP-seq reveals unique and redundant functions in the mature mouse retina. *PLoS One* **9**, e89110 (2014).
11. J. C. Corbo *et al*., CRX ChIP-seq reveals the *cis*-regulatory architecture of mouse photoreceptors. *Genome Res.* **20**, 1512–1525 (2010).
12. RetNet Retinal Information Network. https://sph.uth.edu/retnet/. Accessed 31 January 2019.
13. D. Y. Nishimura *et al*., The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nat. Genet.* **19**, 140–147 (1998).
14. C. L. Freund *et al*., Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* **91**, 543–553 (1997).
15. T. Jordan *et al*., The human PAX6 gene is mutated in two patients with aniridia. *Nat. Genet.* **1**, 328–332 (1992).
16. N. M. Ghiasvand *et al*., Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nat. Neurosci.* **14**, 578–586 (2011).
17. S. Bhatia *et al*., Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* **93**, 1126–1134 (2013).
18. J. Nathans *et al*., Molecular genetics of human blue cone monochromacy. *Science* **245**, 831–838 (1989).
19. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
20. E. P. Consortium; ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
21. M. P. Creyghton *et al*., Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).
22. A. Rada-Iglesias *et al*., A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
23. T. K. Kim *et al*., Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
24. C. Y. McLean *et al*., GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
25. D. Baylor, How photons start vision. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 560–565 (1996).
26. R. H. Steinberg, R. A. Linsenmeier, E. R. Griff, Three light-evoked responses of the retinal pigment epithelium. *Vision Res.* **23**, 1315–1323 (1983).
27. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
28. A. Kanda, J. S. Friedman, K. M. Nishiguchi, A. Swaroop, Retinopathy mutations in the bZIP protein NRL alter phosphorylation and transcriptional activity. *Hum. Mutat.* **28**, 589–598 (2007).
29. S. Wang, C. Sengel, M. M. Emerson, C. L. Cepko, A gene regulatory network controls the binary fate decision of rod and bipolar cells in the vertebrate retina. *Dev. Cell* **30**, 513–527 (2014).
30. S. G. Landt *et al*., ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
31. P. A. Ruzycki, X. Zhang, S. Chen, CRX directs photoreceptor differentiation by accelerating chromatin remodeling at specific target sites. *Epigenet. Chromatin* **11**, 42 (2018).
32. A. E. Hughes, J. M. Enright, C. A. Myers, S. Q. Shen, J. C. Corbo, Cell type-specific epigenomic analysis reveals a uniquely closed chromatin architecture in mouse rod photoreceptors. *Sci. Rep.* **7**, 43184 (2017).
33. M. A. White *et al*., A simple grammar defines activating and repressing *cis*-regulatory elements in photoreceptors. *Cell Rep.* **17**, 1247–1254 (2016).
34. F. Spitz, E. E. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
35. I. Aldiri *et al*., The dynamic epigenetic landscape of the retina during development, reprogramming, and tumorigenesis. *Neuron* **94**, 550–568.e10 (2017).
36. E. Z. Macosko *et al*., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
37. A. M. Klein *et al*., Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
38. B. S. Clark *et al*., Single-cell RNA-seq analysis of retinal development identifies NFI factors as regulating mitotic exit and late-born cell specification. *Neuron* **102**, 1111–1126.e5 (2019).
39. M. Menon *et al*., Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat. Commun.* **10**, 4902 (2019).
40. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
41. T. Stuart *et al*., Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
42. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
43. K. M. Bumsted O'Brien *et al*., Expression of photoreceptor-specific nuclear receptor NR2E3 in rod photoreceptors of fetal human retina. *Invest. Ophthalmol. Vis. Sci.* **45**, 2807–2812 (2004).
44. S. K. John, J. E. Smith, G. D. Aguirre, A. H. Milam, Loss of cone molecular markers in rhodopsin-mutant human retinas with retinitis pigmentosa. *Mol. Vis.* **6**, 204–215 (2000).
45. F. Wu, D. Sapkota, R. Li, X. Mu, Onecut 1 and Onecut 2 are potential regulators of mouse retinal development. *J. Comp. Neurol.* **520**, 952–969 (2012).
46. E. Ferda Percin *et al*., Human microphthalmia associated with mutations in the retinal homeobox gene CHX10. *Nat. Genet.* **25**, 397–401 (2000).
47. J. Gong *et al*., Distribution of vesicular glutamate transporters in rat and human retina. *Brain Res.* **1082**, 73–85 (2006).
48. C. J. Jeon, E. Strettoi, R. H. Masland, The major cell populations of the mouse retina. *J. Neurosci.* **18**, 8936–8946 (1998).
49. J. Carroll *et al*., Deletion of the X-linked opsin gene array locus control region (LCR) results in disruption of the cone mosaic. *Vision Res.* **50**, 1989–1999 (2010).
50. D. S. Kim, T. Matsuda, C. L. Cepko, A core paired-type and POU homeodomain-containing transcription factor program drives retinal bipolar cell gene expression. *J. Neurosci.* **28**, 7748–7764 (2008).
51. F. Haeseleer *et al*., Five members of a novel Ca^{2+}-binding protein (CABP) subfamily with similarity to calmodulin. *J. Biol. Chem.* **275**, 1247–1260 (2000).
52. C. Zeitz *et al*., Mutations in CABP4, the gene encoding the Ca^{2+}-binding protein 4, cause autosomal recessive night blindness. *Am. J. Hum. Genet.* **79**, 657–667 (2006).
53. M. S. Wilken *et al*., DNase I hypersensitivity analysis of the mouse brain and retina identifies region-specific regulatory elements. *Epigenet. Chromatin* **8**, 8 (2015).
54. M. J. Hollenberg, A. W. Spira, Human retinal development: Ultrastructure of the outer retina. *Am. J. Anat.* **137**, 357–385 (1973).
55. A. W. Spira, M. J. Hollenberg, Human retinal development: Ultrastructure of the inner retinal layers. *Dev. Biol.* **31**, 1–21 (1973).
56. Y. F. S. Lu *et al*., Single-cell analysis of human retina identifies evolutionarily conserved and species-specific mechanisms controlling development. bioRxiv:10.1101/779694 (2 October 2019).
57. R. Madelaine *et al*., A screen for deeply conserved non-coding GWAS SNPs uncovers a MIR-9-2 functional mutation associated to retinal vasculature defects in human. *Nucleic Acids Res.* **46**, 3517–3531 (2018).
58. T. S. Scerri *et al*.; MacTel Project Consortium, Genome-wide analyses identify common variants associated with macular telangiectasia type 2. *Nat. Genet.* **49**, 559–567 (2017).
59. P. J. Gordon *et al*., Lhx2 balances progenitor maintenance with neurogenic output and promotes competence state progression in the developing retina. *J. Neurosci.* **33**, 12197–12207 (2013).
60. S. Q. Shen, E. Turro, J. C. Corbo, Hybrid mice reveal parent-of-origin and Cis- and *trans*-regulatory effects in the retina. *PLoS One* **9**, e109382 (2014).
61. M. Lek *et al*.; Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
62. L. A. Lettice *et al*., A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
63. A. R. Janecke *et al*., Mutations in RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy. *Nat. Genet.* **36**, 850–854 (2004).
64. D. A. Thompson *et al*., Retinal degeneration associated with RDH12 mutations results from decreased 11-*cis* retinal synthesis due to disruption of the visual cycle. *Hum. Mol. Genet.* **14**, 3865–3875 (2005).
65. D. Hartl, A. R. Krebs, J. Jüttner, B. Roska, D. Schübeler, *cis*-Regulatory landscapes of four cell types of the retina. *Nucleic Acids Res.* **45**, 11607–11621 (2017).
66. D. P. Murphy, A. E. Hughes, K. A. Lawrence, C. A. Myers, J. C. Corbo, *cis*-Regulatory basis of sister cell type divergence in the vertebrate retina. *eLife* **8**, e48216 (2019).
67. J. Wang *et al*., ATAC-seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat. Commun.* **9**, 1364 (2018).
68. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
69. S. Hrvatin *et al*., Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).
70. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
71. Y. Zhang *et al*., Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
72. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).