

ARTICLE

<https://doi.org/10.1038/s42003-019-0430-6>

OPEN

Genetic analyses of human fetal retinal pigment epithelium gene expression suggest ocular disease mechanisms

Boxiang Liu^{1,6}, Melissa A. Calton^{2,6}, Nathan S. Abell², Gillie Benchorin², Michael J. Gloudemans³, Ming Chen², Jane Hu⁴, Xin Li⁵, Brunilda Balliu⁵, Dean Bok⁴, Stephen B. Montgomery^{2,5} & Douglas Vollrath²

The retinal pigment epithelium (RPE) serves vital roles in ocular development and retinal homeostasis but has limited representation in large-scale functional genomics datasets. Understanding how common human genetic variants affect RPE gene expression could elucidate the sources of phenotypic variability in selected monogenic ocular diseases and pinpoint causal genes at genome-wide association study (GWAS) loci. We interrogated the genetics of gene expression of cultured human fetal RPE (fRPE) cells under two metabolic conditions and discovered hundreds of shared or condition-specific expression or splice quantitative trait loci (e/sQTLs). Co-localizations of fRPE e/sQTLs with age-related macular degeneration (AMD) and myopia GWAS data suggest new candidate genes, and mechanisms by which a common *RDH5* allele contributes to both increased AMD risk and decreased myopia risk. Our study highlights the unique transcriptomic characteristics of fRPE and provides a resource to connect e/sQTLs in a critical ocular cell type to monogenic and complex eye disorders.

¹Department of Biology, Stanford University, Stanford, CA 94305, USA. ²Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. ³Program in Biomedical Informatics, Stanford University School of Medicine, Stanford 94305 CA, USA. ⁴Department of Ophthalmology, Jules Stein Eye Institute, UCLA, Los Angeles 90095 CA, USA. ⁵Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁶These authors contributed equally: Boxiang Liu, Melissa A. Calton. Correspondence and requests for materials should be addressed to S.B. M. (email: smontgom@stanford.edu) or to D.V. (email: vollrath@stanford.edu)

The importance of vision to humans and the accessibility of the eye to examination have motivated the characterization of more than one thousand genetic conditions involving ocular phenotypes¹. Among these, numerous monogenic diseases exhibit considerable inter-familial and intra-familial phenotypic variability^{2–7}. Imbalance in allelic expression of a handful of causative genes has been documented⁸, but few common genetic variants responsible for such effects have been discovered.

Complementing our knowledge of numerous monogenic ocular disorders, recent genome-wide association studies (GWAS)⁹ have identified hundreds of independent loci associated with polygenic ocular phenotypes such as age-related macular degeneration (AMD), the leading cause of blindness in elderly individuals in developed countries^{10,11}, and myopia, the most common type of refractive error worldwide and an increasingly common cause of blindness^{12–14}. Despite the rapid success of GWAS in mapping novel ocular disease susceptibility loci, the functional mechanisms underlying these associations are often obscure.

Connecting changes in molecular functions such as gene expression and splicing with specific GWAS genomic variants has aided the elucidation of functional mechanisms. Non-coding variants account for a preponderance of the most significant GWAS loci^{15,16}, and most expression quantitative trait loci (eQTLs) map to non-coding variants¹⁷. Thousands of eQTLs have been found in a variety of human tissues¹⁸, but ocular cell-types are underrepresented among eQTL maps across diverse tissues.

The retinal pigment epithelium (RPE) is critical for eye development¹⁹ and for an array of homeostatic functions essential for photoreceptors²⁰. Variants of RPE-expressed genes have been associated with both monogenic and polygenic ocular phenotypes, including AMD and myopia. We recently implicated an eQTL associated with an RPE-expressed gene as modulating the severity of inherited photoreceptor degeneration in mice²¹.

To investigate the potential effects of genetically encoded common variation on human RPE gene expression, we set out to identify eQTLs and splice quantitative trait loci (sQTLs) for human fetal RPE (fRPE) cells cultured under two metabolic conditions. Here we describe hundreds of loci of each type, some of which are condition-specific, and connect the mitochondrial oxidation of glutamine with increased expression of lipid synthesis genes, a pathway important in AMD. We find that common variants near genes with disproportionately high fRPE expression explain a larger fraction of risk for both AMD and myopia than variants near genes enriched in non-ocular tissues. We show that a particular variant in *RDH5* is associated with increased skipping of a coding exon, nonsense-mediated decay (NMD) of the aberrant transcript, and three-fold lower minor allele-specific expression. The e/sQTL marked by this variant localizes with high statistical significance with GWAS loci for both AMD and myopia risk, but with opposing directions of effect. Our study lays a foundation for linking e/sQTLs in a critical ocular cell type to mechanisms underlying monogenic and polygenic eye diseases.

Results

The transcriptome of human fRPE cells. We studied 23 primary human fRPE lines (Supplementary Data 1), all generated by the same method in a single laboratory²² and cultured for at least 10 weeks under conditions that promote a differentiated phenotype²³. DNA from each line was genotyped at 2,372,784 variants. Additional variants were imputed and phased using Beagle v4.1²⁴ against 1000 Genomes Phase 3²⁵ for a total of ~13 million variants after filtering and quality control (see Methods section). Comparison of fRPE chromosome 1 genotypes to those of 104 samples from 1000 Genomes indicated that our cohort is

mostly African American in origin, with 4 samples of European ancestry (Supplementary Fig. 1).

Our goal was to identify RPE eQTLs relevant to the tissue's role in both developmental and chronic eye diseases. The balance between glycolytic and oxidative cellular energy metabolism changes during development and differentiation²⁶, and loss of RPE mitochondrial oxidative phosphorylation capacity may contribute to the pathogenesis of AMD²⁷, among other mechanisms. We therefore obtained transcriptional profiles of each fRPE line cultured in medium that favors glycolysis (glucose plus glutamine) and in medium that promotes oxidative phosphorylation (galactose plus glutamine)²⁸. We performed 75-base paired-end sequencing to a median depth of 52.7 million reads (interquartile range: 45.5 to 60.1 million reads) using a paired sample design to minimize batch effects in differential expression analysis (Supplementary Data 2). To determine the relationship between primary fRPE and other tissues, we visualized fRPE in the context of 53 tissues from the GTEx Project v7¹⁸. The fRPE samples formed a distinct cluster situated between heart and skeletal muscle and brain (Fig. 1a), tissues that, like the RPE, are metabolically active and capable of robust oxidative phosphorylation.

To identify genes with disproportionately high levels of expression in the fRPE, we compared the median reads per kilobase of transcript per million mapped reads (RPKM) of fRPE genes against GTEx tissues. We defined fRPE-selective genes as those with median expression at least four standard deviations above the mean (see Methods section). Under this definition, we found 100 protein-coding genes and 30 long non-coding RNAs (lncRNAs) to be fRPE-selective (Fig. 1b and Supplementary Data 3). Multiple previously defined RPE “signature” genes^{29–31} are present in our list including *RPE65* (Fig. 1c) and *RGR* (Fig. 1d). Using this set of genes, we performed Gene Set Enrichment Analysis (GSEA)³² against 5,917 gene ontology (GO) annotations³³. The two gene sets most enriched with fRPE-selective genes were *pigment granule* and *sensory perception of light stimulus* ($FDR < 1 \times 10^{-3}$), consistent with the capacity of fRPE to produce melanin and the tissue's essential role in the visual cycle. Supplementary Data 4 lists the 29 GO pathways enriched using a conservative FWER < 0.05. Recurrent terms in enriched pathway annotations such as pigmentation, light, vitamin, protein translation, endoplasmic reticulum and cellular energy metabolism suggest specific functions that are central to fRPE and outer retinal homeostasis.

Transcriptomic differences across two metabolic conditions.

To gain insight into the response of fRPE cells to altered energy metabolism, we compared gene expression between the two culture conditions using DESeq2³⁴, correcting for sex, ancestry, RIN, and batch (see Methods section). A total of 837 protein coding and lncRNA genes showed evidence of significant differential expression ($FDR < 1 \times 10^{-3}$, Fig. 2a and Supplementary Data 5). Notably, three of the top ten differentially expressed genes are involved in lipid metabolism (*SCD*, *INSIG1*, and *HMGCS2* in order). *SCD* codes for a key enzyme in fatty acid metabolism³⁵, and its expression in RPE is regulated by retinoic acid³⁶. *INSIG1* encodes an insulin-induced protein that regulates cellular cholesterol concentration³⁷. *HMGCS2* encodes a mitochondrial enzyme that catalyzes the first step of ketogenesis³⁸, and this enzyme plays a crucial role in phagocytosis-dependent ketogenesis in fRPE³⁹. To understand the broader impact induced by changes in energy metabolism, we performed pathway enrichment analysis using GSEA³² and found that the top two up-regulated pathways in galactose medium are cholesterol homeostasis and mTORC1 signaling ($FDR < 1 \times 10^{-4}$, Fig. 2b).

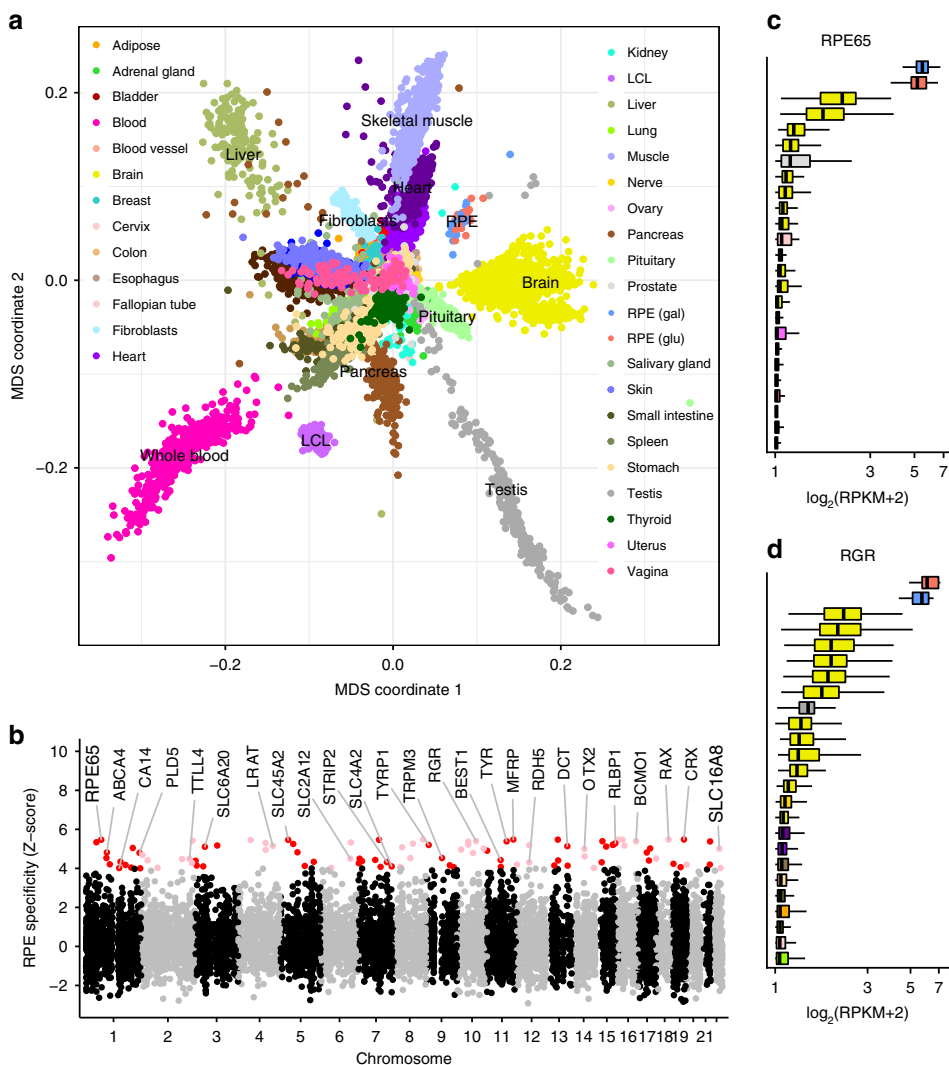


Fig. 1 Characteristics of the fRPE transcriptome. **a** Multidimensional scaling against GTEx tissues locates fRPE near heart, skeletal muscle, and brain samples. **b** A subset of the fRPE-selective gene set defined by z-score >4 is shown including RPE signature genes such as *RPE65* and new genes such as *TYR*. Red/pink dots indicate fRPE-selective genes with z-score >4 in both glucose and galactose conditions. **c, d** Two examples of the expression levels of fRPE-selective genes in various GTEx tissues. Only the top 25 tissues are plotted for visual clarity. For **a, c, and d**, red indicates fRPE glucose condition and blue indicates fRPE galactose condition. For **c** and **d**, each element of the boxplot is defined as follows: centerline, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range

Consistent with the cholesterol finding, forcing cells to rely primarily on oxidation of glutamine for ATP generation increases expression of a suite of genes that promotes lipid synthesis and import (Fig. 2c).

fRPE-selective genes are enriched in genetic ocular diseases.

Disease-associated genes can have elevated expression levels in effector tissues⁴⁰. To determine whether ocular disease genes have elevated expression levels in fRPE, we used a manually curated list of 257 ocular disease-related genes⁴¹ (see Methods section). Compared to all other protein-coding genes, ocular disease-related genes are more specific to fRPE (two-sided *t*-test *p*-value: 1.6×10^{-10}). Further, ocular disease gene expression demonstrated a higher specificity to fRPE than to GTEx tissues (Fig. 3a), suggesting fRPE as a model system for a number of eye diseases. As a control, we repeated the analysis for epilepsy genes (*n* = 189) and observed elevated expression levels in brain tissues as we expected (Supplementary Fig. 2).

Unlike Mendelian ocular diseases, polygenic ocular disorders are characterized by variants with smaller effect sizes scattered

throughout the genome. Using two well-powered GWAS of AMD⁴² and myopia⁴³, we performed stratified linkage disequilibrium (LD) score regression to determine the heritability explained by fRPE. Using a previously established pipeline⁴⁴, we selected the top 500 tissue-enriched genes for fRPE and various GTEx tissues and assigned variants within one kilobase of these genes to each tissue (see Methods section). Risk variants for both AMD and myopia were more enriched around fRPE-selective genes than GTEx tissue-selective genes (Fig. 3b, c). As an assessment of the robustness of the LD score regression results, we repeated the analysis with the top 200 and 1000 tissue-specific genes. A high ranking for fRPE was consistent across all three cutoffs (Supplementary Fig. 3).

e/sQTL discovery. To determine the genetic effects on gene expression in fRPE, we used RASQUAL⁴⁵ to map eQTLs by leveraging both gene-level and allele-specific count information to boost discovery power. Multiple-hypothesis testing for both glucose and galactose conditions was conducted jointly with a hierarchical procedure called TreeQTL⁴⁶. At FDR < 0.05, we

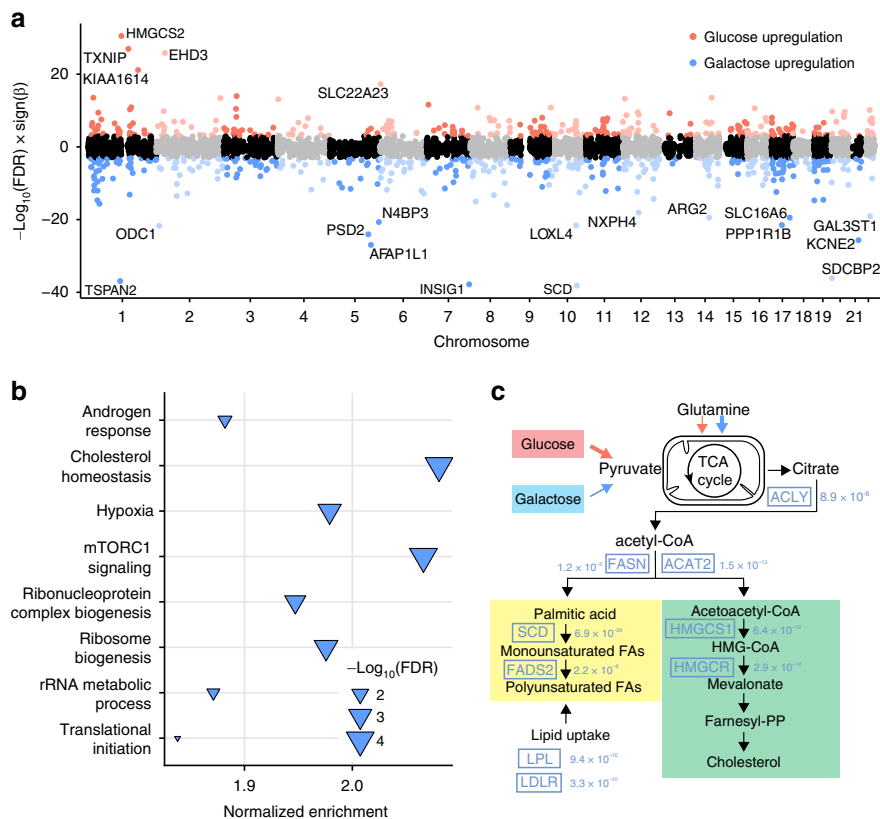


Fig. 2 Differential expression across two metabolic conditions. **a** Transcriptome-wide differential expression patterns: red indicates upregulated in glucose, blue indicates upregulated in galactose. **b** Gene set enrichment analysis of differentially expressed genes. The pathway most enriched is cholesterol homeostasis (upregulated in galactose condition). **c** Key genes involved in cholesterol biosynthesis and import are upregulated in response to the increased oxidation of glutamine that occurs in the galactose condition. Estimated FDR values are shown next to the gene names

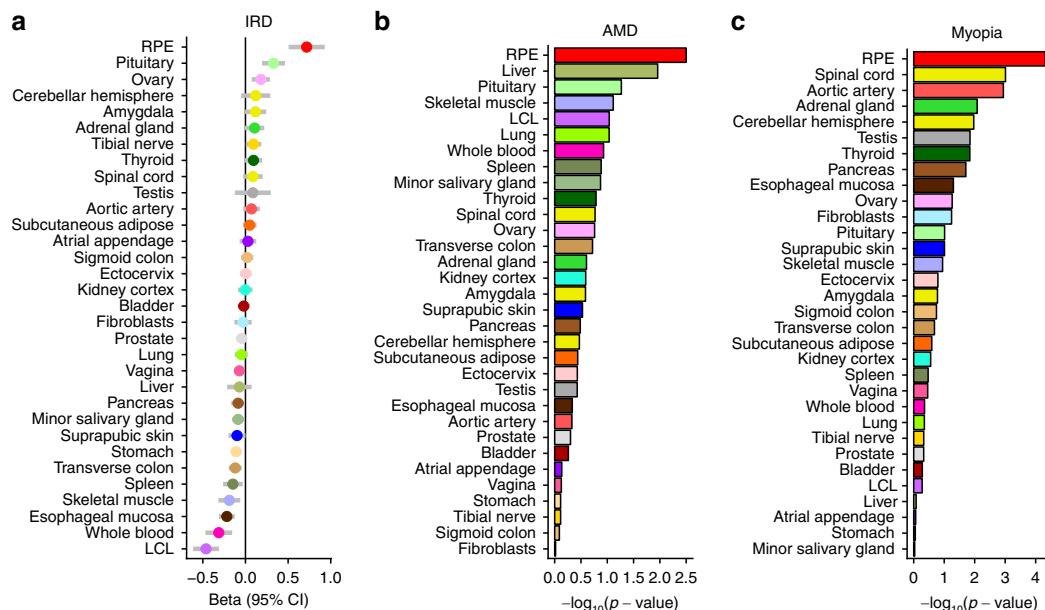


Fig. 3 fRPE-selective genes are enriched in monogenic and polygenic diseases. **a** Genes causal for inherited retinal disorders (IRD) have elevated expression in fRPE. **b, c** Variants near RPE-selective genes explain a larger proportion of AMD (**b**) and myopia (**c**) risk than those near GTEx tissue-selective genes. The red bar represents the top 500 fRPE-selective genes

found 687 shared, 264 glucose-specific, and 166 galactose-specific eQTLs (Table 1, Supplementary Data 6 and 7, Fig. 4a and Supplementary Figs. 4 and 5). An example of a shared eQTL is *RGR* (Fig. 4d), which encodes a G protein-coupled receptor that is mutated in retinitis pigmentosa⁴⁷. An example of a glucose-specific eQTL is *ABCA1* (Fig. 4b), which encodes an ATP-binding cassette transporter that regulates cellular cholesterol efflux⁴⁸. Common variants near *ABCA1* have been associated with glaucoma⁴⁹ and AMD⁴². An example of a galactose-specific eQTL is *PRPF8* (Fig. 4c), which encodes a splicing factor⁵⁰. *PRPF8* mutations are a cause of autosomal dominant retinitis pigmentosa⁵¹ and lead to RPE dysfunction in a mouse model⁵².

Differential expression alone is unlikely to account for the condition-specific nature of the eQTLs we identified because only about a quarter are differentially expressed (FDR < 0.05) and almost all of these exhibit an absolute fold change of less than two. Rather, it is likely that regulatory specificity is the underlying cause of these eQTLs. We therefore used HOMER⁵³ to identify transcription factor binding motifs enriched around metabolic-specific eQTLs (see Methods section). Two motifs, TEAD1 ($p < 1 \times 10^{-6}$) and ZEB1 ($p < 1 \times 10^{-3}$), are among the top five motifs in the galactose condition (Supplementary Data 8). TEAD1 is known to play a role in aerobic glycolysis reprogramming⁵⁴, and ZEB1 is known to render cells resistant to glucose deprivation⁵⁵. We did not find enriched motifs in the glucose condition for transcription factors with well-known metabolic functions.

We compared fRPE to GTEx eGenes using a previously established two-step FDR approach⁵⁶. We used fRPE-shared eGenes (FDR < 0.05 in both metabolic conditions) as the discovery set to remove any treatment-dependent regulatory effect, and used GTEx eGenes with a relaxed threshold (FDR < 0.1) as the replication set. eGenes from the discovery set not recapitulated in the replicated set were defined as fRPE-selective eGenes. This approach returned three genes (Fig. 4e–g): *TYR*, encoding an oxidase controlling the production of melanin; *CRX*, encoding a transcription factor critical for photoreceptor differentiation; and *MFRP*, encoding a secreted WNT ligand important for eye development. The *TYR* eQTL maps to a variant (rs4547091) previously described as located in an OTX2 binding site and responsible for modulating *TYR* promoter activity in cultured RPE cells⁵⁷. All three genes are also fRPE-selective genes (Fig. 1), suggesting that apparent regulatory specificity is the by-product of expression selectivity. We also compared our eGenes to the EyeGEx database⁵⁸. Among the 687 eGenes shared across both conditions, 498 (72.5%) are also eGenes reported in EyeGEx.

We also assessed the genetic effect on splicing by quantifying intron usages with LeafCutter⁵⁹ and mapping splicing quantitative trait loci (sQTL) with FastQTL⁶⁰ in permutation mode to obtain intron-level p -values. Following an established approach⁵⁹, we used a conservative Bonferroni correction across introns within each intron cluster and calculated FDR across cluster-level p -values (see Methods section). We found 210 and 193 sQTLs at FDR < 0.05 for glucose and galactose conditions, respectively (Table 2, Supplementary Data 9 and 10). The top sQTL in the glucose condition regulates splicing in *ALDH3A2* (FDR < 2.06×10^{-9}), which codes for an aldehyde dehydrogenase isozyme involved in lipid metabolism⁶¹. Mutations in this gene cause Sjogren-Larsson syndrome⁶², which can affect the macular RPE⁶³. The top sQTL in the galactose condition regulates splicing of transcripts encoding *CAST*, a calcium-dependent protease inhibitor involved in the turnover of amyloid precursor protein⁶⁴.

Fine mapping of complex ocular disease risk loci. To assess whether specific instances of GWAS signals can be explained by eQTL or sQTL signals, we performed colocalization analysis with

Table 1 Expression QTL discoveries

Trait type	No. tested	No. of eQTLs (FDR < 0.05)		
		Glucose	Galactose	Shared
Protein coding	12,515	254	163	656
lincRNA	582	10	3	31
Total	13,097	264	166	687

a modified version of eCAVIAR⁶⁵ (see Methods section). All variants within a 500-kilobase window around any GWAS (p -value < 1×10^{-4}) or QTL (p -value < 1×10^{-5}) signal were used as input to eCAVIAR, and any locus with colocalization posterior probability (CLPP) > 0.01 was considered significant. To identify condition-specific colocalization events, we ran eCAVIAR separately for two metabolic conditions (see Methods section). For the AMD GWAS, we identified four eQTL colocalization events for each condition (Supplementary Fig. 6). One of these, *WDR5*, demonstrates glucose-specific colocalization (CLPP: glucose = 0.033 and galactose = 0.002). For the myopia GWAS, we identified three and seven colocalization events for galactose and glucose conditions, respectively. Three are condition specific (*PDE3A*, *ETS2*, and *ENTPD5*; Supplementary Fig. 7). For example, *PDE3A*, shows galactose-specific colocalization (CLPP: glucose = 0.0004; galactose = 0.014). eQTLs at *PARP12* and *CLU* colocalized with AMD and myopia signals, respectively, under both conditions (Fig. 5a, d, Supplementary Figs. 6 and 7). While neither locus reached genome-wide significance in the respective GWAS, the significant co-localizations we describe implicate *PARP12* and *CLU* as new candidate genes for these disorders.

Among the four genes exceeding our threshold for eQTL and AMD GWAS colocalization, *RDH5*, encoding a retinol dehydrogenase that catalyzes the conversion of 11-cis retinol to 11-cis retinal in the visual cycle⁶⁶, showed the most significant signal (Fig. 5a and Supplementary Data 11). *RHD5* was previously suggested as an AMD candidate gene⁴², but no mechanism was proposed. Two tightly linked AMD-associated variants (rs3138141 and rs3138142, $r^2 = 0.98$) are highly correlated with *RDH5* expression (Fig. 5b). The minor haplotype identified by the rs3138141 “a” allele is associated with a significantly smaller percentage of total *RDH5* expression (26.4%) than the major haplotype identified by the “C” allele (73.6%) (Fig. 5g). We found no evidence for an effect on transcripts from the adjacent *BLOC1S1* gene or on *BLOC1S1*-*RDH5* read-through transcripts. The same variants mark an *RDH5* sQTL (Fig. 5a, c) associated with differences in the usage of exon 3 of the transcript; samples that are heterozygous at rs3138141 (Ca) exhibit an average of more than three times the amount of exon 3 skipping compared to CC homozygous samples (Fig. 5h and Supplementary Fig. 8). The same e/sQTL also colocalized with a myopia GWAS signal (Fig. 5d–f, Supplementary Data 12), suggesting a mechanism for the prior association of the *RDH5* locus with myopia⁶⁷ and refractive error¹³.

NMD as a putative mechanism underlying an *RDH5* eQTL.

The association of the rs3138141/2 minor haplotype with both an *RDH5* eQTL and sQTL suggests a mechanistic relationship. We estimate that ~80% of isoforms transcribed from the “C” haplotype are normal, whereas ~75% of isoforms transcribed from the “a” haplotype are mis-spliced (see Methods section). The increased skipping of exon 3 (out of 5) associated with the minor haplotype results in more transcripts with a frameshift and a premature termination codon (PTC) near the 5' end of exon 4. Many mammalian transcripts with PTCs are subject to NMD, particularly when the PTC is not located in the last exon⁶⁸.

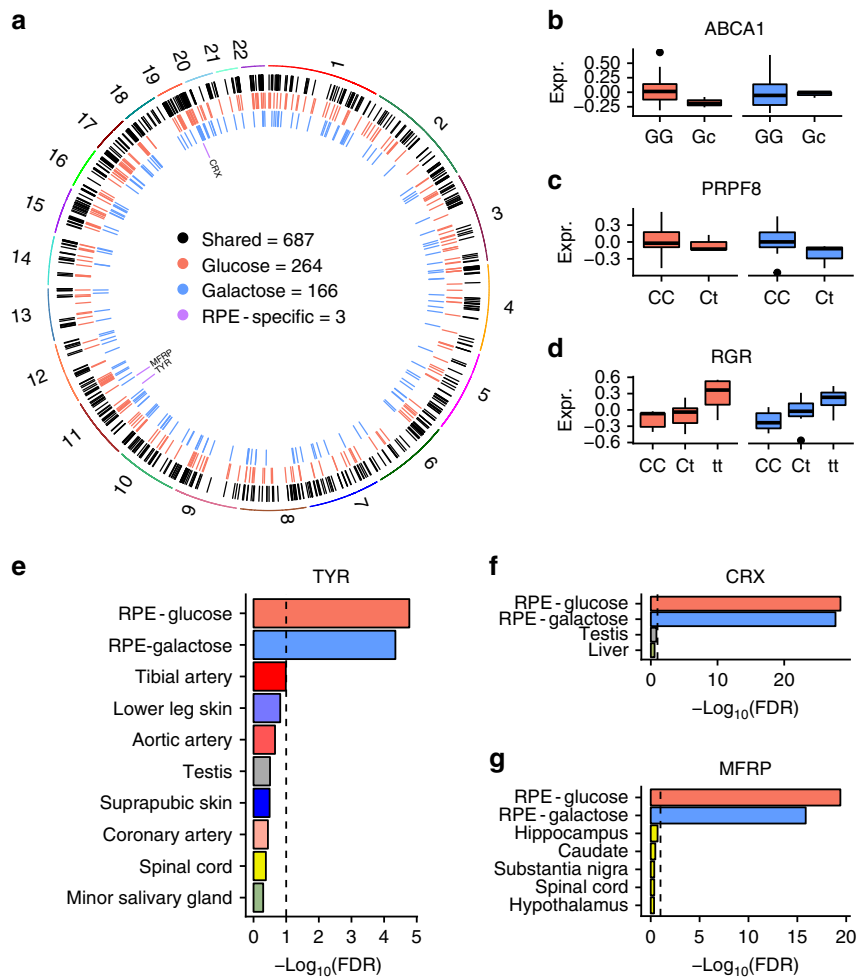


Fig. 4 Landscape of genetic regulation of RPE gene expression. **a** We discovered 687, 264, and 166 eQTLs that are shared, glucose-specific, and galactose-specific, respectively. Comparison with GTEx eGenes revealed three shared eGenes that are currently unique to frPE. **b** A glucose-specific eQTL in *ABCA1*. **c** A galactose-specific eQTL in *PRPF8*. **d** A shared eQTL in *RGR*. The y-axis of panels **b–d** denotes normalized expression values. **e–g** Evidence for frPE-specificity for three eQTLs compared to GTEx. Black dashed lines indicate FDR = 0.1. Minor alleles are indicated by lowercase. For **b, c,** and **d,** each element of the boxplot is defined as follows: centerline, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range

Table 2 Splicing QTL discoveries

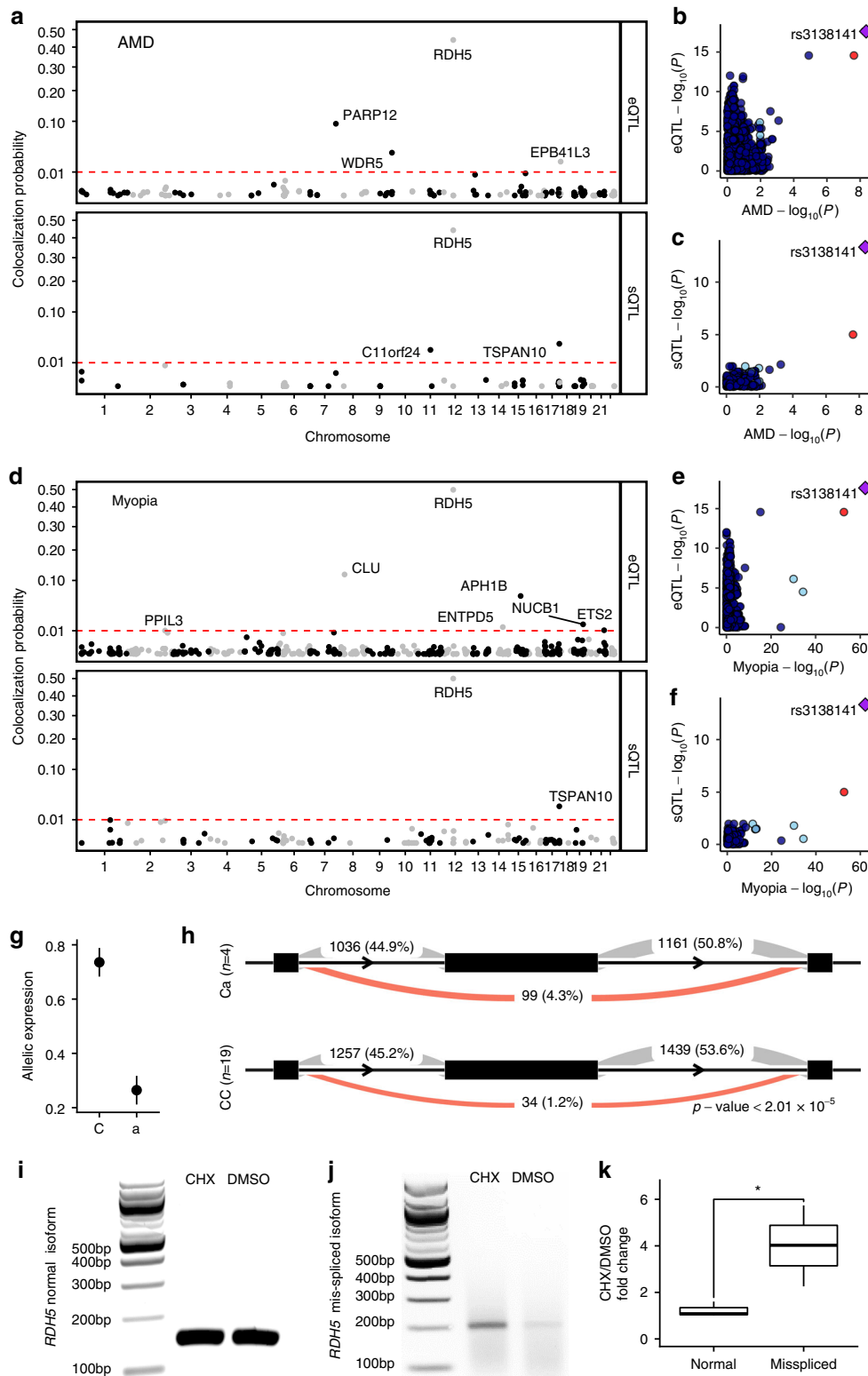
Condition	Trait type	No. tested	No. of sQTLs		
			FDR < 0.05	FDR < 0.01	FDR < 0.001
Glucose	Protein coding	15,516	205	129	66
	lncRNA	187	5	4	4
	Total	15,703	210	133	70
Galactose	Protein coding	16,239	188	112	56
	lncRNA	196	5	4	2
	Total	16,435	193	116	58

Treatment of cells with protein synthesis inhibitors such as cycloheximide (CHX) has been shown to increase the abundance of transcripts subject to NMD⁶⁹. To assess a possible role for NMD in the stability of RDH5 transcripts, we treated differentiated immortalized human RPE cells (ARPE-19) with CHX and quantified the abundance of the normal and skipped exon 3 isoforms by RT-PCR. CHX caused a significant increase in the abundance of the skipped exon 3 isoform as compared to the normal (Fig. 5i–k and Supplementary Fig. 9). These data are

consistent with a model in which the minor allele promotes the formation of an aberrant RDH5 mRNA that is subject to NMD, leading to an overall reduction in the steady state levels of RDH5 transcripts.

Discussion

The importance of the RPE for development and lifelong homeostasis of the eye has motivated numerous studies of the RPE transcriptome. Several of these studies proposed similar sets of RPE “signature” genes, the largest of which comprises 171 genes^{29–31}. Only 23 of these genes are present among our group of 100 frPE-selective protein-coding genes. Our approach of comparing frPE expression levels to GTEx data, which almost exclusively derive from adult autopsy tissue specimens, may have captured genes highly expressed in cultured and/or fetal cells. Absence in GTEx of pure populations of specialized cell types, especially ocular, may explain other genes in our set. Still, many of the genes we identified are known to serve vital functions in the RPE, as demonstrated by pathway enrichment for pigment synthesis and visual processes. We also identified 30 enriched lncRNAs, a class of transcripts not included in previous signature gene sets. The most highly expressed lncRNA in our list, *RMRP*,



is critical for proper mitochondrial DNA replication and OXPHOS complex assembly in HeLa cells⁷⁰, but its role in the RPE has not yet been investigated. RPE-enriched genes whose functions have not been studied in the tissue afford opportunities for advancing understanding of this important epithelial layer.

Our findings have potential implications for phenotypic variability in monogenic ocular diseases. Mutations in all three of the fRPE-selective eGenes cause monogenic eye diseases. For

example, heterozygous mutations in the transcription factor *CRX* cause dominant forms of photoreceptor degeneration, which can exhibit variable age at onset and disease progression among members of the same family⁴. Genetically encoded variation in the transcript levels of normal or mutant *CRX* alleles may contribute to such variable expressivity. Indeed, mouse models of *CRX*-associated retinopathies provide evidence for a threshold effect in which small changes in expression cause large differences

Fig. 5 Fine mapping of disease-associated variants using fRPE gene regulation. **a** Colocalization posterior probability for fRPE e/sQTLs with AMD. **b, c** Scatter plots demonstrate clear colocalization between AMD GWAS signal at rs3138141 and *RDH5* eQTL (**b**) and sQTL (**c**). **d** Colocalization posterior probability for fRPE e/sQTLs with myopia. **e, f** Scatter plots demonstrate clear colocalization between myopia GWAS signal at rs3138141, the same variant identified for AMD, and *RDH5* eQTL (**e**) and sQTL (**f**). **a–f** Colocalization results are with glucose QTLs. Galactose QTL colocalizations can be found in Figs. S18–19. **g** Relative allelic expression estimated by RASQUAL with 95% confidence intervals is shown. **h** Increased skipping of *RDH5* exon 3 (middle black rectangle) is associated with the minor allele at rs3138141. The average read counts are shown for three splice junctions in groups of fRPE cells with different genotypes. The proportion of counts for all three sites for a given junction and genotype is shown in parenthesis. Exon and intron lengths are not drawn to scale. Minor alleles are indicated by lowercase. **i** Gel image showing *RHD5* normal isoform amplified from CHX or DMSO treated ARPE-19 cells. **j** Gel image showing *RHD5* mis-spliced isoform amplified from CHX or DMSO treated ARPE-19 cells. **k** Relative fold change between CHX and DMSO treatments for normal and mis-spliced RNA isoforms. Error bars indicate standard error of the mean for $n = 3$ independent experiments. * $p < 0.05$

in phenotype⁷¹. Mutations in *MFRP* cause extreme hyperopia (farsightedness). Affected individuals usually have two mutant alleles, but inheritance of a lower-expressing normal allele could explain an affected heterozygous individual in a family with otherwise recessive disease⁵. The substantial number of fRPE eQTLs associated with other ocular diseases (Fig. 3a) supports a contribution of common genetic variants to the widespread phenotypic variability observed in monogenic eye disorders.

Our findings also have implications for complex ocular diseases. Evidence suggests that defects in RPE energy metabolism contribute to the pathogenesis of AMD, the hallmark of which is accumulation of cholesterol rich deposits in and around the RPE^{72,73}. Forcing fRPE cells to rely on oxidation of glutamine, the most abundant free amino acid in blood, caused upregulation of genes involved in the synthesis of cholesterol, monounsaturated and polyunsaturated fatty acids, as well as genes associated with lipid import. Transcripts for three of the upregulated genes (*FADS1*, *FADS2*, and *ACAT2*) are increased in macular but not extramacular RPE from individuals with early-stage AMD⁷⁴.

Co-localization of the same *RDH5* e/sQTL with both AMD and myopia GWAS loci suggests risk mechanisms for these very different complex diseases. The rs3138141/2 minor haplotype confers an elevated risk for AMD⁴², but is protective for myopia^{13,43,67}. Reduction in *RDH5* activity as a risk factor for AMD is consistent with rare *RDH5* loss-of-function mutations that cause recessive fundus albipunctatus, which can include macular atrophy^{75,76}. More puzzling is the relationship between lower *RDH5* transcript levels (and presumably enzyme activity) and a reduced risk of myopia. *RDH5* is best known for its role in the regeneration of 11-cis retinal in the visual cycle, but the enzyme has also been reported to be capable of producing retinoids suitable for retinoic acid signaling^{77,78}. Evidence from animal models implicates retinoic acid in eye growth regulation¹², and retinal all-trans retinoic acid levels are elevated in a guinea pig model of myopia⁷⁹. Thus the same allele, which has risen to substantial frequencies in some populations (0.38 minor allele frequency in South Asians and 0.19 in Europeans <https://www.ncbi.nlm.nih.gov/projects/SNP/>), may dampen retinoic acid signaling during eye development and growth, and later contribute to chronic photoreceptor dysfunction in older adults.

The eye is a highly specialized organ with limited representation in large-scale functional genomics datasets. Our analysis of genetic variation and metabolic processes in fRPE cells, even with modest sample sizes, expands our ability to map functional variants with potential to contribute to complex and monogenic eye diseases. Future studies with larger sample sizes from geographically diverse populations, and/or targeting other ocular cell types, will likely discover additional e/sQTLs and functional variants involved in genetic eye diseases.

Methods

Sample acquisition and cell culture. Primary human fetal RPE (fRPE) lines were isolated from fetal eyes (Advanced Biosciences Resources, Inc., Alameda, CA) by collecting and freezing non-adherent cells cultured in low calcium medium as

described²². When needed, fRPE cells were thawed and plated onto 6-well plates in medium as described²³ with 15% FBS. The next day, medium was changed to 5% FBS and the cells were allowed to recover for two additional days. Cells were then trypsinized in 0.25% Trypsin-EDTA (Life Technologies Corporation), resuspended in medium with 15% FBS and plated onto human extracellular matrix-coated (BD Biosciences) Corning 12-well transwells (Corning Inc., Corning, NY) at 240 K cells per transwell. The next day medium was changed to 5% FBS. Cells were cultured for at least 10 weeks to become differentiated (transepithelial resistance of $>200 \Omega \times \text{cm}^2$) and highly pigmented. Medium with 5% FBS was changed every 2–3 days. For the galactose and glucose specific culture conditions, differentiated fRPE cells were cultured for 24 h prior to RNA isolation in DMEM medium (Sigma) with 1 mM sodium pyruvate (Sigma), 4 mM L-glutamine (Life Technologies Corporation), 1% Penicillin-Streptomycin (Life Technologies Corporation), and either 10 mM D-(+)-glucose (Sigma) or 10 mM D-(+)-galactose (Sigma)²⁸. The fRPE lines studied here are not available for distribution.

Genotype data and quality control

Microarray library preparation and genotyping. All 24 RPE samples were genotyped on three Illumina Infinium Omni2.5-8 BeadChip using the Infinium LCG Assay workflow (<https://www.illumina.com/products/by-type/microarray-kits/infinium-omni25-8.html>). A total of 200 ng of genomic DNA was extracted and amplified to generate sufficient quantity of each individual DNA sample. The amplified DNA samples were fragmented and hybridized overnight on the Omni2.5-8 BeadChip. The loaded BeadChips went through single-base extension and staining, and were imaged on the iScan machine to obtain genotyping information. Genotyping data were exported from Illumina GenomeStudio to ped and map pairs, merged, and converted to the VCF format using PLINK v1.9⁸⁰. We removed variants that were missing in more than 5% of samples.

Variant annotation. We annotated variants using genomic features (including downstream-gene variant, exonic variant, intronic variant, missense variant, splice-acceptor variant, splice-donor variant, splice-region variant, synonymous variant, upstream-gene variant, 3'-UTR variant, 5'-UTR variant), loss-of-function, and nonsense-mediated decay predictions, and clinical databases (including ClinVar, OMIM and OrphanNet) using SnpEff v4.31⁸¹.

Imputation and phasing. We used Beagle v4.1⁸² to perform genotype imputation and phasing. Genotypes were imputed and phased with 1000 Genomes Project phase 3 reference panel. Before imputation and phasing, we filtered the original VCF file to only bi-allelic SNP sites on autosomes and removed sites with more than 5% missing genotypes. We also re-coded the VCF file based on the reference and alternative allele designation of 1000 Genomes Project phase 3 reference panel using the conform-gt program which was provided with the Beagle software.

Quality control. Prior to imputation, we performed standard pre-imputation QC by removing variants that are missing in more than 5% of samples and used the filtered call set as input to Beagle. After imputation and phasing, we removed variants with allelic $r^2 < 0.8$ (higher allelic r^2 indicates higher confidence). Standard post-imputation QC also requires removing variants with low Hardy-Weinberg Equilibrium (HWE) p -value. Due to the extensive admixture in our study cohort, we reasoned that HWE may have trouble in distinguishing genotyping error from admixture. We opted to not apply HWE but instead to remove multi-allelic variants (variants with more than two alleles). The Ts/Tv ratio of the filtered call set was above 2.0 for all chromosomes but chromosome 8 (=1.98) and 16 (=1.93). Several chromosomes have Ts/Tv ratio greater than 2.1, indicating that they are enriched with known (vs. novel) variants (Supplementary Fig. 10). To detect sample duplication, we plotted genotype correlation across every pair of samples and found two samples (sample 3 and 5) were duplicates of each other (Supplementary Fig. 11). We removed one sample (sample 5) at random. The filtered VCF files were used for downstream analysis.

Sex determination. We determined biological sex of each donor using genotyping information. We extracted the genotype dosage with bcftools and calculated the proportion of heterozygous SNPs (heterozygous: dosage = 1; homozygous: dosage = 0 or 2) for chromosome 1 and X. Donors with low heterozygosity on

chromosome X (proportion of heterozygous ≈ 0) SNP were defined as males. Chromosome 1 was used as control to establish the baseline for heterozygosity. This cohort has 11 male and 13 female individuals (Supplementary Fig. 12).

Ancestry determination. We determined ancestry of each donor using genotype information. We extracted genotype dosages of chromosome 1 for 23 RPE samples and 4 samples from each of the 26 populations (for a total of 104 individuals) in 1000 Genomes phase 3 version 5 dataset²⁵. We calculated the principal components using the `prcomp` function in R. The top three principal components explained the most variability (Supplementary Fig. 13), and were used for downstream analysis. The first two principal components clearly separates the European, African, and Asian populations. Four RPE samples are European, and the rest are admixed. Most admixed individuals are African American (Supplementary Fig. 1).

Transcriptomic data and quality control

RNA-seq library preparation and sequencing. RNA was extracted using TRIzol Reagent (Invitrogen) per manufacturer instructions. RNA sequencing was performed on all samples with an RNA integrity number (RIN) of 8.0 or higher and with at least 500 ng total RNA. Stranded, poly-A+ selected RNA-seq libraries were generated using the Illumina TruSeq Stranded mRNA protocol. We performed 75 bp paired-end RNA sequencing on an Illumina NextSeq 500 on all RPE samples (Supplementary Data 2). Glucose and galactose samples from each line were sequenced together to minimize batch effects.

RNA sequencing read mapping. Raw data was de-multiplexed using `bcl2fastq2` from Illumina with default parameters. Reads were aligned against the hg19 human reference genome with STAR (v2.4.2a)⁸³ using GENCODE v19 annotations⁸⁴ and otherwise default parameters. After alignment, duplicate reads were marked using Picard MarkDuplicates (v2.0.1) and reads marked as duplicates or with non-perfect mapping qualities were removed.

Gene and splicing event quantification. We used HTSeq v0.6.0⁸⁵ to count the number of reads overlapping each gene based on the GENCODE v19 annotation. We counted reads on the reverse strand (ideal for Illumina's TruSeq library), required a minimum alignment quality of 10, but otherwise used default parameters. We also quantified RPKM using RNA-SeQC v1.1.8⁸⁶ using hg19 reference genome and GENCODE v19 annotation with flags “-noDoC -strictMode” but otherwise default parameters. We quantified allele-specific expression using the `createASVCF.sh` script from RASQUAL⁴⁵ with default parameters. For splicing quantification, we used LeafCutter⁵⁹ to determine intron excision levels with default parameters. Briefly, we first converted bam files to splice junction counts (`bam2junc.sh`) and clustered introns based on sharing of splice donor or acceptor sites (`leafcutter_cluster.py`). For each cluster, we required a minimum number of 30 reads, and a minimum fraction of 0.1% in support of each junction. We required each intron must not exceed 100 kbp.

Quality control. We profiled the RNA-seq library to a median depth of 52.7 million reads (interquartile range: 45.5–60.1 million reads), for a total of 2.5 billion reads (Supplementary Fig. 14a). We checked the number of uniquely mapped reads to ensure sufficient number of mapped reads. The RNA-seq libraries have a median number of 46.8 million (88.8%) uniquely mapped reads, with an interquartile range of 41.0–55.2 million reads (Supplementary Fig. 14b). We ran `VerifyBamID`⁸⁷ with parameters “-ignoreRG—best” on RNA-seq BAM files using genotype VCF files as reference and did not find any sample swaps.

Normalization of quantifications. We extract hidden factors from RNA sequencing data using surrogate variable analysis (sva)⁸⁸ jointly (protecting the treatment variable) and separately for glucose and galactose-treated samples. Prior to estimating hidden factors, the raw count gene expression data was library size corrected, variance stabilized, and log₂ transformed using the R package DESeq2³⁴. Genes with average read count below 10 and with zero counts in more than 20% of samples were considered not expressed and filtered (to remove tails). A total of 15,056 and 15,062 expressed genes remained for glucose and galactose, respectively. Since library size correct depends on all genes, filtered genes were again corrected for library size, variance stabilized, and log₂ transformed using DESeq2 before being used as input to sva. We ran sva, as implemented in the sva R package, with default parameters, and obtained seven significant surrogate variables with joint analysis and four and five significant surrogate variables for the glucose and galactose condition, respectively. We also extracted surrogate variables for splicing level quantification. The joint, glucose, and galactose analysis returned four, two, and two factors, respectively.

Correlation between known and hidden confounders. We calculated the correlation between known (treatment, RIN, sequencing batch, sex, and ancestry) and hidden (surrogate variables) factors to determine which factors to include in downstream analyses. The jointly inferred factors (seven in total) captures treatment (factor 5, $r = 0.91$), RIN (factor 1, $r = 0.71$) and batch effect (factor 7, $r = -0.56$), but does not capture sex (best $r = -0.18$) or ancestry (best $r = -0.34$). This agrees with the intuition that treatment, RIN and batch effect have broad influences on gene

expression measurement, while sex and ancestry only influence a small set of relevant genes (Supplementary Fig. 15a). To reduce the correlation between factor 5 and treatment, we ran supervised sva protecting the treatment effect. Even with protection on treatment, factor 5 remains correlated ($r = -0.62$), likely due to the strong and broad effect exerted by metabolic perturbation (Supplementary Fig. 15b). The glucose surrogate variables captured RIN (factor 1, $r = 0.81$) and batch (factor 2, $r = 0.78$), and the galactose surrogate variables captured RIN (factor 1, $r = -0.61$) and batch (factor 1 and 4, $r = -0.54$ and -0.5 , respectively). None of the glucose or galactose surrogate variables captured sex or ancestry (Supplementary Figs. 15c, d). We also compared surrogate variables from splicing quantification. Without protecting the treatment, surrogate variable 3 from the joint set correlated with the treatment (Supplementary Fig. 16a). Even after protecting the treatment, surrogate variable 1 from the joint set correlated with the treatment (Supplementary Fig. 16b), similar to what was observed for expression surrogate variables. Surrogate variables from the glucose and galactose condition correlated strongly with RIN ($r = -0.74$ and -0.76 , respectively, Supplementary Fig. 16c, d).

External datasets

RNA-seq and eQTL datasets. We used GTEx V7¹⁸ as a reference dataset to perform RPE-selective gene and RPE-specific eQTL analyses. The GTEx V7 dataset collected 53 tissues across 714 donors. All tissues across all donors were used in RPE-selective gene analysis. Among the 53 tissues, 48 tissues have sufficient sample size to perform eQTL analysis and were used for RPE-specific eQTL calling.

GWAS datasets. We used two well-powered ocular disorder GWAS datasets to perform colocalization analyses. The AMD study⁴² is a meta-analysis across 26 studies and identified 52 independent GWAS signals, including 16 novel loci. The myopia GWAS was part of a 42-trait GWAS collection aimed at finding shared genetic influences across different traits⁴³.

Ocular disease genes dataset. We used ocular disease genes from the Genetic Eye Disease (GEDi) test panel⁴¹, which encompasses 257 genes in total including known inherited retinal disorder genes (IRD, $n = 214$), glaucoma and optic atrophy genes ($n = 8$), candidate IRD genes ($n = 24$), age-related macular degeneration risk genes ($n = 9$), and a non-syndromic hearing loss gene ($n = 1$).

RPE-selective gene and pathway enrichment analyses

Expression z-score method. To identify RPE-selective genes (high expression in RPE relative to other tissues), we inferred expression specificity using the following procedure.

1. Calculate the median expression level (x) across all individuals for each tissue.
2. Calculate the mean (μ) and standard deviation (σ) of median expression values across tissues.
3. Derive a z-score for each tissue as follows: $z = (x - \mu) / \sigma$.
4. Define a gene to be tissue-selective if its z-score is greater than 4.

We filtered out genes on sex and mitochondrial chromosomes, and further filtered out genes in HLA region due to low mappability. To determine whether technical confounders (such as batch effect) affected RPE z-scores, we used a QQ-plot to visualize the z-score of each tissue against the average z-score across tissues. To calculate the average z-scores, we ranked genes within each tissue and take the average z-score for genes with the same rank across tissues. The average z-scores represent the expected distribution. If the z-score distribution from a tissue is markedly different from the expected distribution, this distribution will separate from the diagonal on the QQ plot. Supplementary Fig. 17 shows that RPE z-scores situate within the midst of z-scores from GTEx tissues. In fact, the only outlier is testis, which is a known outlier from previous studies.

Pathway enrichment of RPE-selective genes. To identify coordinated actions by RPE-selective transcriptomic elements, we performed GSEA³² using z-scores as input against GO gene sets from the Molecular Signature Database⁸⁹ with 10,000 permutations and otherwise default parameters. The full results are in Supplementary Data 4.

Differential expression and pathway enrichment analyses

Differential expression analysis. We performed differential expression analysis with DESeq2³⁴ to detect genes whose expression levels were affected by metabolic perturbation. Due to the correlation between hidden factors (SVs) and treatment, we decided to use known factors in the DESeq2 model. More specifically, we observed moderate correlation between SVs and condition ($r = -0.62$) even after protecting for condition (Supplementary Fig. 15). Adding SVs that are correlated with conditions will bias the estimates on the coefficient for conditions. The model is shown below:

$$E(\text{expression}) = \beta_0 + \beta_1 \cdot \text{treatment} + \beta_2 \cdot \text{sex} + \sum_{i=1}^3 \beta_{a_i} \cdot \text{PC}_i + \beta_r \cdot \text{RIN} + \beta_b \cdot \text{batch}$$

Pathway enrichment analysis. We performed pathway enrichment analysis with GSEA using a ranked gene list with 10,000 permutations but otherwise default

parameters. The ranking metric was calculated by multiplying the $-\log_{10}(\text{FDR})$ by the sign of the effect size from DESeq2. For the pathway database, we used a subset of the Molecular Signatures Database composed of its Hallmark, Biocarta, Reactome, KEGG and GO gene sets⁸⁹.

fRPE-selective genes in ocular diseases

RPE-selective expression in ocular disease genes. We stratified all protein-coding genes into two groups: (1) ocular disease genes ($n = 257$) and (2) non-ocular disease genes ($n = 18,477$). To determine whether known ocular disease genes have elevated expression in fRPE, we compared the expression specificity z -score distribution (defined previously) across these two groups with a two-sided t -test. We performed the same analysis for all GTEx tissues as a benchmark. As a control, we repeated this analysis using known epilepsy genes ($n = 189$) curated from the Invitae epilepsy gene test panel (<https://www.invitae.com/en/physician/tests/03401/>).

RPE-selective expression in ocular disease GWA studies. GWAS risk loci are frequently enriched around causal genes, which have elevated expression in relevant tissues⁹⁰. To determine whether variants around RPE-selective genes explain higher disease heritability than expected by chance, we performed stratified LD score regression on tissue-selective genes using a previously established pipeline⁴⁴. Since LD score regression operates on a variant level, we assigned variants within 1-kb around any exon of tissue-selective genes to each tissue. Although many variants show long-range interaction, we restricted our analysis to a conservative window size to capture only nearby cis-effects. We performed LD score regression on the 200, 500, and 1000 tissue-specific genes (Fig. 3 and Supplementary Fig. 3).

eQTL mapping and quality control

Covariate selection. We determined biological sex, genomic ancestry, and hidden confounders as described in previous sections. We performed covariate selection by empirically maximizing the power to detect eQTL. We randomly selected 50 genes from chromosome 22 to perform covariate selection for computational feasibility and to avoid overfitting. We added sex, genotype principal components (maximum of three), and surrogate variables sequentially. We chose not to include batch effect or RIN because they were well represented by surrogate variables. We tested the top three genotype principal components because they explained most of the variability in the genotyping data (Supplementary Fig. 13). After multiple hypothesis correction, the number of eAssociations (defined as a SNP-gene pair that passed hierarchical multiple hypothesis testing by TreeQTL⁴⁶) increased monotonically for both glucose and galactose conditions as the number of covariates increased (Supplementary Fig. 18), which agrees with our intuition that sva only returns significant and independent surrogate variables. Therefore, we decided to use sex, top three genotype principal components and all surrogate variables (four and five for glucose and galactose conditions, respectively).

Per-treatment eQTL calling. We mapped eQTL using RASQUAL⁴⁵, which integrates total read count with allele-specific expression (ASE) to boost power for eQTL mapping. To obtain GC-corrected library size, we first calculated GC content using GENCODE v19⁸⁴ by taking the average GC content of all exons of a given gene. Next, we calculated GC-corrected library sizes were calculated based on read count output from HTSeq v0.6.0⁸⁵. We used sex, ancestry principal components, and all surrogate variables. Mathematically, the model is the following:

$$E(\text{expression}) = \beta_0 + \beta_g \cdot \text{genotype} + \beta_s \cdot \text{sex} + \sum_{i=1}^3 \beta_{a,i} \cdot \text{PC} + \sum_{i=1}^n \beta_{s,i} \cdot \text{SV}$$

where e stands for expressions, g stands for genotypes, PC stands for genotype principal components, SV stands for surrogate variables, and $n = 4$ and 5 for glucose and galactose conditions, respectively. We obtained gene-level and association-level FDR using a hierarchical hypothesis correction procedure implemented in TreeQTL⁴⁶. TreeQTL uses a hierarchical FDR correction procedure, which performs FDR correction first on the gene level, and then on the association level (gene by SNP). We used FDR < 0.05 on both gene and association levels.

eQTL quality control. We determined whether the p -values were inflated (e.g., due to model mis-specification) by visualizing their distribution. The distribution suggests that the p -values are slightly conservative. The spike around zero and upward trend in the QQ plot shows clear enrichment for significant eQTL (Supplementary Fig. 19a, b). As expected, eQTLs with low p -values were enriched around transcription start sites (Supplementary Fig. 19c, d).

Differential eQTL calling with TreeQTL. We performed multi-tissue eQTL calling, using the RASQUAL p -values and the multi-tissue version of TreeQTL⁴⁶. We set the gene as the first level, the treatment as the second level, and the gene-treatment-SNP as the third level and used the default FDR < 0.05 cutoff for all three levels. We showed a comparison of $-\log_{10}(p\text{-value})$ across two metabolic conditions in Supplementary Fig. 4, in which the top five treatment-specific and shared eQTL are labeled. We ranked the differential eQTL result in the order of decreasing $\delta|\pi - 0.5|$ (difference in allelic imbalance in two conditions). More specifically, π denotes

the allelic ratio (alternative allele/reference allele), and $|\pi - 0.5|$ denotes the allelic imbalance. The difference in allelic imbalance, $\delta|\pi - 0.5|$, defines the change in eQTL effect size across two conditions. We show the allelic ratios in Supplementary Fig. 5.

RPE-selective eQTL. We compared RPE and GTEx eGenes with a two-step FDR approach as described previously⁵⁶. In brief, eGenes shared across both conditions in fRPE were selected (FDR < 0.05). We decided to filter for shared eGenes because they likely reflect regulatory effects not due to treatments. For each eGene, we screened all GTEx tissues for association at a relaxed FDR < 0.1 and defined an eGene as RPE-selective if no significant association were found in GTEx. Note that such strategy is conservative on two levels. First, by selecting shared eQTL in RPE, these eQTL must pass FDR < 0.05 in both treatments. Second, GTEx FDR corrections were performed tissue-by-tissue, and per-tissue FDR is anti-conservative. We also compared fRPE eGenes to retinal eGenes. The EyeGEx dataset used 406 samples to map eQTLs in the retina and found a total of 10,463 eGenes. We again selected eGenes found in both glucose and galactose conditions ($n = 687$) and grouped them into fRPE-specific and EyeGEx-shared if they were also eGenes in the EyeGEx dataset.

Motif enrichment in treatment-specific eQTLs. In order to find motifs enriched around treatment-specific eQTLs, we first selected the lead eQTLs from either condition and extracted the 15 bp flanking the lead SNP as the target sequences. To obtain matched background sequence, we flipped the eQTL SNP to its alternative allele as the background. To keep the direction of effect consistent, we always used the expression-increasing allele as the target and the expression-decreasing allele as the background. The target and background sequences were used as input to HOMER to identify enriched motifs.

sQTL calling and quality control

Covariate selection. We performed covariate selection by empirically maximizing the power to detect sQTL. We used intron clusters only from chromosome 1 to avoid overfitting, and tested only the top three genotype principal components because they explained most of the variability in the genotyping data (Supplementary Fig. 13). FastQTL were run in permutation mode (adaptively permute 100–10,000 times) to obtain intron-level sQTL p -values. After multiple hypothesis correction, the number of significant sQTL cluster decreased as the number of covariates increased (Supplementary Fig. 20). This is likely because LeafCutter uses the ratio between each intron and its intron cluster as the phenotype. Suppose a batch effect influences the expression of a gene. Such batch effect will influence the quantification of each intron in the same direction. Taking the ratio between a intron and its intron cluster effectively cancels out the batch effect.

Per-treatment sQTL calling. We mapped sQTLs separately for two conditions using FastQTL⁶⁰ in both nominal and permutation modes and used a simple linear regression:

$$E(\text{intron}) = \beta_0 + \beta_g \cdot \text{genotype}$$

where s stands for the ratio between reads overlapping each intron and the total number of reads over-lapping the intron cluster, g stands for genotypes. To obtain cluster-level p -values, we used a conservative approach to correct for family-wise error rate with the Bonferroni procedure across introns within each cluster. Global FDR estimates were calculated using the lowest Bonferroni adjusted p -values per cluster. We used FDR < 0.05 as a significance cutoff.

sQTL quality control. As a quality control, we determined whether the p -values were inflated by visualizing their distribution. The p -values showed a uniform distribution with a spike near 0 (Supplementary Fig. 21a). The upward trend in the QQ plot shows clear enrichment for significant eQTL (Supplementary Fig. 21b). Further, sQTLs with low p -values were enriched around splicing donor and acceptor sites (Supplementary Fig. 21c), and intronic sQTL SNPs were enriched at intron boundaries (Supplementary Fig. 21d).

Fine-mapping of polygenic ocular disease risk loci. We used fRPE eQTL and sQTL information to identify potential causal genes in two well-powered GWAS on age-related macular degeneration and myopia using a modified version of eCAVIAR⁶⁵. For every significant eQTL, we tested all variants within 500-kb of the lead eQTL SNP for colocalization with GWAS summary statistics. At each candidate locus, we ran FINEMAP⁹¹ twice to compute the posterior probability that each individual SNP at the locus was a causal SNP for the GWAS phenotype and fRPE e/sQTLs. We then processed the FINEMAP results to compute a colocalization posterior probability (CLPP) using the method described by eCAVIAR⁶⁵. We defined any locus with CLPP > 0.01 to have sufficient evidence for colocalization. At loci that showed colocalization between RPE eQTLs and GWAS associations, we performed the colocalization tests again using eQTLs from each of 44 GTEx tissues. To determine whether any potential causal genes act primarily through fRPE, we repeated colocalization analysis with GTEx eQTLs (Supplementary Figs. 22 and 23). To identify condition-specific colocalization, we ran eCAVIAR separately for the glucose and galactose conditions. A condition-specific colocalization is defined

as having a CLPP > 0.01 in one condition, and at least an order of magnitude lower in the other condition with a CLPP < 0.01.

Estimation of isoform proportions. To estimate the proportions of the normal and mis-spliced isoform (exon-3 skipped isoform), we solved a system of equations based on the following observations:

The “C” haplotype produces approximately 3 times as much as the “a” haplotype.

1. The mis-spliced isoform accounts for approximately 1% of expression in individuals with CC genotype.
2. The mis-spliced isoform accounts for approximately 4% of expression in individuals with Ca genotype.

We use n_c and n_a to denote the proportion of normal isoform for the “C” and the “a” haplotypes, and p_n and p_m to denote the proportion of normal and mis-spliced isoforms that pass non-sense mediated decay (not degraded). For simplicity, we assume that $p_n = 1$ because normal isoform should not be degraded by NMD. We use c_t , c_n , and c_m to denote the total, normal, and mis-spliced isoforms for the “C” haplotype, and a_t , a_n , and a_m to denote the total, normal, and mis-spliced isoforms for the “a” haplotype. We know that:

$$c_n = 100c_m$$

$$c_n + a_n = 25(c_m + a_m)$$

$$c_t = 3a_t$$

Plugging in $c_n = n_c$, $c_m = (1 - n_c)p_m$, $a_n = n_a$ and $a_m = (1 - n_a)p_m$:

$$n_c = 100(1 - n_c)p_m$$

$$n_c + n_a = 25((1 - n_c)p_m + (1 - n_a)p_m)$$

$$n_c + (1 - n_c)p_m = 3(n_a + (1 - n_a)p_m)$$

Solving the system of equations leads to $n_c = 0.82$, $n_a = 0.25$, and $p_m = 0.05$. In other words, 82% and 25% of isoforms transcribed from the “C” and “a” haplotypes are normal, respectively. We estimate that NMD will degrade 95% of mis-spliced isoforms.

Experimental validation. ARPE-19 cells were obtained from ATCC (CRL-2302). The cells were obtained directly from ATCC within the past year. They exhibit the expected cobblestone morphology and slight pigmentation when differentiated by standard protocols. ARPE-19 cells were fixed, stained with DAPI and imaged by fluorescence microscopy. No evidence of mycoplasma contamination was seen. ARPE-19 cells were differentiated for 3 months in 6-well plates (Corning) in medium containing 3 mM pyruvate⁹² and treated with 100 µg/mL cycloheximide (CHX; Sigma) or vehicle (DMSO) for 3 h. Cells were then collected, RNA was extracted by TRIzol (Invitrogen), and cDNA was synthesized with an iScript™ cDNA Synthesis Kit (Bio-RAD). Oligonucleotide primers were designed to specifically amplify the normal or mis-spliced isoforms of the *RDH5* transcript. For the normal isoform, the forward primer (ggggctactgtctcctcaaa) was located in exon 3 and the reverse primer (tgccaggtttttccagct) was located in exon 4, with an expected product size of 151 bp. The amplification conditions were: 94 °C 2 min followed by 38 cycles of 94 °C 30 s, 60 °C 30 s, 72 °C 15 s. For the mis-spliced isoform, the forward primer (gatgcagcttaaggaagcag/gcg) spanned the exon 2/4 junction with the three bases at the 3' end located in exon 4. The reverse primer (gcgctgttcatttcaggt) was located in exon 5. The expected product size is 204 bp. The amplification conditions were: 94 °C 2 min followed by 50 cycles of 94 °C 30 s, 60 °C 30 s, 72 °C 15 s. AmpliTaq (ThermoFisher) and 2.5 mM MgCl₂ were used for all reactions. The identities of the normal and mis-spliced PCR products were confirmed by Sanger sequencing. For quantification, PCR products were resolved on 2% agarose gels containing ethidium bromide and imaged using a Bio-Rad ChemiDoc Touch Imaging System. Equal-sized boxes were drawn around bands for the CHX and DMSO samples, grayscale values were measured by ImageJ (NIH), and the relative fold change was calculated (mean ± SEM; three independent experiments). A one-sided Student's *t*-test was used to assess the statistical significance of a model under which CHX increased product abundance.

Statistics and reproducibility. To promote the reproducibility of our study, we deposited raw experimental data to GEO (see Data availability) and open sourced all scripts for data processing and analysis (see Code availability).

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

Data availability

All relevant data are available in the Supplementary Data files (Supplementary Data 1–12). Full eQTL and sQTL summary statistics have been made deposited into Box: <https://stanford.box.com/s/asrxy066xxe1j7mfj56p3z3d405gijj> and are available at <http://>

montgomerylab.stanford.edu/resources.html. RNAseq data can be downloaded via GEO accession number GSE129479. Source data underlying the figures is available as Supplementary Data 13–30.

Code availability

Code to reproduce all analyses in this manuscript has been deposited on GitHub: <https://github.com/boxiangliu/rpe>.

Received: 13 October 2018 Accepted: 17 April 2019

Published online: 20 May 2019

References

1. McKusick, V. A. Mendelian inheritance in man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
2. Boon, C. J. F. et al. The spectrum of retinal dystrophies caused by mutations in the peripherin/RDS gene. *Prog. Retin. Eye Res.* **27**, 213–235 (2008).
3. Nash, B. M., Wright, D. C., Grigg, J. R., Bennetts, B. & Jamieson, R. V. Retinal dystrophies, genomic applications in diagnosis and prospects for therapy. *Transl. Pediatr.* **4**, 139–163 (2015).
4. Paunescu, K., Preising, M. N., Janke, B., Wissinger, B. & Lorenz, B. Genotype–phenotype correlation in a German family with a novel complex CRX mutation extending the open reading frame. *Ophthalmology* **114**, 1348–1357.e1341 (2007).
5. Sundin, O. H. et al. Extreme hyperopia is the result of null mutations in MFRP, which encodes a Frizzled-related protein. *Proc. Natl Acad. Sci.* **102**, 9553–9558 (2005).
6. Vaclavik, V., Gaillard, M. C., Tiab, L., Schorderet, D. F. & Munier, F. L. Variable phenotypic expressivity in a Swiss family with autosomal dominant retinitis pigmentosa due to a T494M mutation in the PRPF3 gene. *Mol. Vis.* **16**, 467–475 (2010).
7. Sergouniotis, P. I. et al. Phenotypic variability in RDH5 retinopathy (Fundus Albipunctatus). *Ophthalmology* **118**, 1661–1670 (2011).
8. Llavona, P. et al. Allelic expression imbalance in the human retinal transcriptome and potential impact on inherited retinal diseases. *Genes* **8**, 283 (2017).
9. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
10. Bressler, N. M. Age-related macular degeneration is the leading cause of blindness. *JAMA* **291**, 1900–1901 (2004).
11. Swaroop, A., Chew, E. Y., Bowes Rickman, C. & Abecasis, G. R. Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration. *Annu. Rev. Genom. Hum. Genet.* **10**, 19–43 (2009).
12. Zhang, Y. & Wildsoet, C. F. RPE and choroid mechanisms underlying ocular growth and myopia. *Prog. Mol. Biol. Transl. Sci.* **134**, 221–240 (2015).
13. 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).
14. Holden, B. A. et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology* **123**, 1036–1042 (2016).
15. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
16. Gusev, A. et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
17. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. B* **368**, 20120362–20120362 (2013).
18. Consortium, G. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
19. Raymond, S. M. & Jackson, I. J. The retinal pigmented epithelium is required for development and maintenance of the mouse neural retina. *Curr. Biol.* **5**, 1286–1295 (1995).
20. Strauss, O. The retinal pigment epithelium in visual function. *Physiol. Rev.* **85**, 845–881 (2005).
21. Vollrath, D. et al. Tyro3 modulates MERTK-associated retinal degeneration. *PLoS Genet.* **11**, e1005723 (2015).
22. Hu, J. & Bok, D. Culture of highly differentiated human retinal pigment epithelium for analysis of the polarized uptake, processing, and secretion of retinoids. *Methods Mol. Biol.* **652**, 55–73 (2010).
23. Maminishkil, A. et al. Confluent monolayers of cultured human fetal retinal pigmented epithelium exhibit morphology and physiology of native tissue. *Invest. Ophthalmol. Vis. Sci.* **47**, 3612–3624 (2006).

24. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).
25. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
26. Folmes, C. D. L., Dzeja, P. P., Nelson, T. J. & Terzic, A. Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell* **11**, 596–606 (2012).
27. Terluk, M. R. et al. Investigating mitochondria as a target for treating age-related macular degeneration. *J. Neurosci.* **35**, 7304–7311 (2015).
28. Gohil, V. M. et al. Nutrient-sensitized screening for drugs that shift energy metabolism from mitochondrial respiration to glycolysis. *Nat. Biotechnol.* **28**, 249–255 (2010).
29. Bennis, A. et al. Comparison of mouse and human retinal pigment epithelium gene expression profiles: potential implications for age-related macular degeneration. *PLoS ONE* **10**, e0141597 (2015).
30. Liao, J.-L. et al. Molecular signature of primary retinal pigment epithelium and stem-cell-derived RPE cells. *Hum. Mol. Genet.* **19**, 4229–4238 (2010).
31. Strunnikova, N. V. et al. Transcriptome analysis and molecular signature of human retinal pigment epithelium. *Hum. Mol. Genet.* **19**, 2468–2486 (2010).
32. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci.* **102**, 15545–15550 (2005).
33. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
35. Paton, C. M. & Ntambi, J. M. Biochemical and physiological function of stearoyl-CoA desaturase. *Am. J. Physiol.* **297**, E28–E37 (2009).
36. Samuel, W. et al. Regulation of stearoyl coenzyme A desaturase expression in human retinal pigment epithelial cells by retinoic acid. *J. Biol. Chem.* **276**, 28744–28750 (2001).
37. Yang, T. et al. Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. *Cell* **110**, 489–500 (2002).
38. Aledo, R. et al. Genetic basis of mitochondrial HMG-CoA synthase deficiency. *Hum. Genet.* **109**, 19–23 (2001).
39. Reyes-Reveles, J. et al. Phagocytosis-dependent ketogenesis in retinal pigment epithelium. *J. Biol. Chem.* **292**, 8038–8047 (2017).
40. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30**, 2496–2497 (2014).
41. Consugar, M. B. et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet. Med.* **17**, 253–261 (2015).
42. 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**, 134–143 (2015).
43. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
44. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
45. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **48**, 206–213 (2016).
46. Peterson, C. B., Bogomolov, M., Benjamini, Y. & Sabatti, C. TreeQTL: hierarchical error control for eQTL findings. *Bioinformatics* **32**, 2556–2558 (2016).
47. Morimura, H., Saindelle-Ribeaud, F., Berson, E. L. & Dryja, T. P. Mutations in RGR, encoding a light-sensitive opsin homologue, in patients with retinitis pigmentosa. *Nat. Genet.* **23**, 393–394 (1999).
48. Schmitz, G. & Langmann, T. Structure, function and regulation of the ABC1 gene product. *Curr. Opin. Lipidol.* **12**, 129 (2001).
49. Chen, Y. et al. Common variants near *ABCA1* and in *PMM2* are associated with primary open-angle glaucoma. *Nat. Genet.* **46**, 1115–1119 (2014).
50. Luo, H. R., Moreau, G. A., Levin, N. & Moore, M. J. The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* **5**, 893–908 (1999).
51. Tanackovic, G. et al. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum. Mol. Genet.* **20**, 2116–2130 (2011).
52. Farkas, M. H. et al. Mutations in pre-mRNA processing factors 3, 8, and 31 cause dysfunction of the retinal pigment epithelium. *Am. J. Pathol.* **184**, 2641–2652 (2014).
53. Heinz, S. 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).
54. Enzo, E. et al. Aerobic glycolysis tunes YAP/TAZ transcriptional activity. *Embo J.* **34**, 1349–1370 (2015).
55. Kanska, J. et al. Glucose deprivation elicits phenotypic plasticity via ZEB1-mediated expression of NNMT. *Oncotarget* **8**, 26200–26220 (2017).
56. Barreiro, L. B. et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proc. Natl Acad. Sci. USA* **109**, 1204–1209 (2012).
57. Reinisalo, M., Putula, J., Mannermaa, E., Urtti, A. & Honkakoski, P. Regulation of the human tyrosinase gene in retinal pigment epithelium cells: the significance of transcription factor orthodenticle homeobox 2 and its polymorphic binding site. *Mol. Vis.* **18**, 38–54 (2012).
58. Ratnapriya, R. et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0351-9> (2019).
59. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
60. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
61. Kelson, T. L., Secor McVoy, J. R. & Rizzo, W. B. Human liver fatty aldehyde dehydrogenase: microsomal localization, purification, and biochemical characterization. *Biochim. et. Biophys. Acta* **1335**, 99–110 (1997).
62. Nakahara, K. et al. The Sjögren-Larsson syndrome gene encodes a hexadecenal dehydrogenase of the sphingosine 1-phosphate degradation pathway. *Mol. Cell* **46**, 461–471 (2012).
63. Nilsson, S. E. & Jagell, S. Lipofuscin and melanin content of the retinal pigment epithelium in a case of Sjögren-Larsson syndrome. *Br. J. Ophthalmol.* **71**, 224–226 (1987).
64. Hanna, R. A., Campbell, R. L. & Davies, P. L. Calcium-bound structure of calpain and its mechanism of inhibition by calpastatin. *Nature* **456**, 409–412 (2008).
65. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
66. Sahu, B. & Maeda, A. Retinol dehydrogenases regulate vitamin A metabolism for visual function. *Nutrients* **8**, 746 (2016).
67. Kiefer, A. K. et al. Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS Genet.* **9**, e1003299 (2013).
68. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci.* **7**, 26 (2017).
69. Carter, M. S. et al. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *J. Biol. Chem.* **270**, 28995–29003 (1995).
70. Noh, J. H. et al. HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev.* **30**, 1224–1239 (2016).
71. Ruzycski, P. A., Tran, N. M., Kolesnikov, A. V., Kefalov, V. J. & Chen, S. Graded gene expression changes determine phenotype severity in mouse models of CRX-associated retinopathies. *Genome Biol.* **16**, 114 (2015).
72. Curcio, C. A. et al. Esterified and unesterified cholesterol in drusen and basal deposits of eyes with age-related maculopathy. *Exp. Eye Res.* **81**, 731–741 (2005).
73. Pikuleva, I. A. & Curcio, C. A. Cholesterol in the retina: the best is yet to come. *Prog. Retin. Eye Res.* **41**, 64–89 (2014).
74. Ashikawa, Y. et al. Potential protective function of the sterol regulatory element binding factor 1-fatty acid desaturase 1/2 axis in early-stage age-related macular degeneration. *Heliyon* **3**, e00266 (2017).
75. Yamamoto, H. et al. Mutations in the gene encoding 11-cis retinol dehydrogenase cause delayed dark adaptation and fundus albipunctatus. *Nat. Genet.* **22**, 188–191 (1999).
76. Yamamoto, H. et al. A novel RDH5 gene mutation in a patient with fundus albipunctatus presenting with macular atrophy and fading white dots. *Am. J. Ophthalmol.* **136**, 572–574 (2003).
77. Dueter, G. Families of retinoid dehydrogenases regulating vitamin A function. *Eur. J. Biochem.* **267**, 4315–4324 (2001).
78. Nadauld, L. D. et al. Dual roles for adenomatous polyposis coli in regulating retinoic acid biosynthesis and Wnt during ocular development. *Proc. Natl Acad. Sci.* **103**, 13409–13414 (2006).
79. McFadden, S. A., Howlett, M. H. C. & Mertz, J. R. Retinoic acid signals the direction of ocular elongation in the guinea pig eye. *Vis. Res.* **44**, 643–653 (2004).
80. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 559 (2015).
81. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
82. Browning, B. L. & Yu, Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* **85**, 847–861 (2009).
83. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, (bts635–621) (2012).
84. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

85. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
86. Deluca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
87. Jun, G. et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).
88. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
89. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
90. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
91. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
92. Samuel, W. et al. Appropriately differentiated ARPE-19 cells regain phenotype and gene expression profiles similar to those of native RPE cells. *Mol. Vis.* **23**, 60–89 (2017).

Acknowledgements

This work was supported by a Stanford Center for Computational, Evolutionary and Human Genomics Predoctoral Fellowship and the National Key R&D Program of China (2016YFD0400800) (to B.L.); T32EY20485 (to M.A.C.); the Edward Mallinckrodt Jr. Foundation and NIH grants R33HL120757, U01HG009431, R01MH101814, and R01HG008150 (to S.B.M.); The Macular Degeneration Research Program of the BrightFocus Foundation, the Foundation Fighting Blindness, and R01EY025790 (to D. V.); and P30EY026877. D.B. is the Dolly Green Professor of Ophthalmology at UCLA. The authors thank 23andMe employees and participants for providing summary statistics for the myopia GWAS and Baidu USA for providing critical support to B.L.

Author contributions

M.A.C., S.B.M. and D.V. conceived and designed experiments. D.B. and J.H. provided critical reagents and expertise. M.A.C., G.B., B.L. and M.C. performed experiments. B.L., N.S.A., B.B., M.J.G., X.L., S.B.M. and D.V. analyzed data. B.L., D.V., S.B.M., M.A.C. and B.B. wrote the paper.

Additional information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s42003-019-0430-6>.

Competing interests: Stephen Montgomery is on the Scientific Advisory Board of Prime Genomics. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019