Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error

Milly S. Tedja, Robert Wojciechowski, Pirro G. Hysy, Nicholas Eriksson, Nicholas A. Furlotte, Virginie J. M. Verhoeven, Adriana I. Iglesias, Magda A. Meester-Smoor, Stuart W. Tompson, Qiao Fan, Anthony P. Khawaja, Ching-Yu Cheng, René Höhn, Kenji Yamashiro, Adam Wenocur, Clare Grazal, Toomas Haller, Andres Metspalu, Juho Wedenoja, Jost B. Jonas, Ya Xing Wang, Jing Xie, Paul Mitchell, Paul J. Foster, Barbara E. K. Klein, Ronald Klein, Andrew D. Paterson, S. Mohsen Hosseini, Rupal L. Shah, Cathy Williams, Yik Ying Teo, Yih Chung Tham, Preeti Gupta, Wanting Zhao, Yuan Shi, Woei-Yuh Saw, E-Shyong Tai, Xue Ling Sim, Jennifer E. Huffman, Ozren Polašek, Caroline Hayward, Goran Bencic, Igor Rudan, James F. Wilson, CREAM, 23andMe Research Team, UK Biobank Eye and Vision Consortium, Peter K. Joshi, Akitaka Tsujikawa, Fumihiko Matsuda, Kristina N. Whisenhunt, Tanja Zeller, Peter J. van der Spek, Roxanna Haak, Hanne Meijers-Heijboer, Elisabeth M. van Leeuwen, Sudha K. Iyengar, Jonathan H. Las, Albert Hofman, Fernando Rivadeneira, André G. Uitterlinden, Johannes R. Vingerling, Terho Lehtimäki, Olli T. Raitakari, Maria Pina Concas, Tae-Hwi Schwantes-An, Robert P. Igo Jr, Gabriel Cuellar-Partida, Nicholas G. Martin, Jamie E. Craig, Norbert Pfeiffer, Olavi Pärssinen, Paul N. Baird, Veronique Vitart, Najaf Amin, Cornelia M. van Duijn, Joan E. Bailey-Wilson, Terri L. Young, Seang-Mei Saw, Dwight Stambolian, Stuart MacGregor, Jeremy A. Guggenheim, Katie M. Williams, Stuart MacGregor, Jeremy A. Guggenheim, Joyce Y. Tung, Christopher J. Hammond and Caroline C. W. Klaver

Refractive errors, including myopia, are the most frequent eye disorders worldwide and an increasingly common cause of blindness. This genome-wide association meta-analysis in 160,420 participants and replication in 95,505 participants increased the number of established independent signals from 37 to 161 and showed high genetic correlation between Europeans and Asians (>0.78). Expression experiments and comprehensive in silico analyses identified retinal cell physiology and light processing as prominent mechanisms, and also identified functional contributions to refractive-error development in all cell types of the neurosensory retina, retinal pigment epithelium, vascular endothelium and extracellular matrix. Newly identified genes implicate novel mechanisms such as rod-and-cone bipolar synaptic neurotransmission, anterior-segment morphology and angiogenesis. Thirty-one loci resided in or near regions transcribing small RNAs, thus suggesting a role for post-transcriptional regulation. Our results support the notion that refractive errors are caused by a light-dependent retina-to-sclera signaling cascade and delineate potential pathobiological molecular drivers.

Refractive errors are common optical aberrations determined by mismatches in the focusing power of the cornea, lens and axial length of the eye. Their distribution worldwide is rapidly shifting toward myopia, or nearsightedness. The myopia boom is particularly prominent in urban East Asia, where up to 95% of 20-year-olds in cities such as Seoul and Singapore have this

A full list of authors and affiliations appears at the end of the paper.
refractive error\(^1\). The prevalence of myopia is also rising throughout Western Europe and the United States, affecting ~50% of young adults in these regions\(^2\). Although refractive errors can be optically corrected, even at moderate values they carry substantial risk of ocular complications with high economic burden\(^3\). One in three individuals with high myopia (~6 diopters or worse) develop irreversible visual impairment or blindness, mostly as a result of myopic macular degeneration, retinal detachment or glaucoma\(^4\). At the other extreme, high hyperopia predisposes individuals to strabismus, amblyopia and angle-closure glaucoma\(^5\).

Refractive errors result from a complex interplay of lifestyle and genetic factors. The most established lifestyle factors for myopia are high education, lack of outdoor exposure and excessive near work\(^6\). Recent research has identified many genetic variants for refractive errors, myopia and axial length\(^7-10\). Two large studies—the International Consortium for Refractive Error and Myopia (CREAM)\(^11\) and the personal genomics company 23andMe, Inc.\(^12\)—have provided the most comprehensive results\(^13\).

Given that only 3.6% of the variance of the refractive-error trait was explained by the identified genetic variants\(^14\), we presumed a high missing heritability. We therefore combined data from CREAM and 23andMe, and expanded the study sample to 160,420 individuals from a mixed-ancestry population with quantitative information on refraction for a genome-wide association study (GWAS) meta-analysis. Index variants were tested for replication in an independent cohort consisting of 95,505 individuals from the UK Biobank. We conducted systematic comparisons to assess differences in genetic inheritance and the distribution of risk variants between Europeans and Asians. Polygenic risk analyses were performed to evaluate the contributions of the identified variants to the risk of myopia and hyperopia. Finally, we integrated expression data and bioinformatics on the identified genes to gain insight into the possible mechanisms underlying the genetic associations.

**Results**

**Susceptibility loci for refractive error.** We performed a GWAS meta-analysis on adult untransformed spherical equivalent (SphE), using summary statistics from 37 studies from CREAM, and on age of diagnosis of myopia (AODM) from two cohorts from 23andMe\(^2;13\) (Supplementary Fig. 1 and Supplementary Table 1a). The analyses were based on ~11 million genetic variants (SNPs, insertions and deletions) genotyped or imputed to the 1000 Genomes Project Phase 1 reference panel (version 3, March 2012 release\(^14\)) that passed extensive quality control (Supplementary Figs. 2–4 and Supplementary Table 1b).

Meta-analyses were conducted in three stages: stage 1, CREAM (European dataset, CREAM-EUR, number of participants (n) = 44,192; Asian dataset, CREAM-ASN, n = 11,935); stage 2, 23andMe (n = 104,293; Methods); stage 3, joint meta-analysis of stages 1 and 2. Because CREAM and 23andMe applied different phenotype measures, we used signed Z scores as the mean per-allele effect size and assigned equal weights to CREAM and 23andMe. We identified 7,967 genome-wide-significant genetic variants clustering in 140 loci (Fig. 1a, Supplementary Figs. 5 and 6, Supplementary Tables 2–5 and Supplementary Data 1 and 2), replicating all 37 previously discovered loci and finding 104 novel loci. We applied genomic control at each stage and checked for population stratification by using linkage disequilibrium (LD)-score regression\(^15\) (stage 1 and 2 inflation factors (GC) <1.1 and LD-score regression intercepts (LDSC\(_{indep}\)) 0.892–1.023; Supplementary Table 6 and Supplementary Figs. 6 and 7). At stage 3, we observed genomic inflation (GC = 1.129; Supplementary Fig. 6), probably because of true polygenicity rather than population stratification or cryptic relatedness\(^16\). LDSC\(_{indep}\) remained undetermined, owing to mixed ancestry.

To detect the presence of multiple independent signals at the discovered loci, a stepwise conditional analysis was performed with GCTA-COJO\(^17\) on summary statistics from all European cohorts (n = 148,485), with the Rotterdam Study I–III (RSI–III) used as a reference panel for LD structure (n\(_{BSI-M}\) = 10,775). This analysis yielded 27 additional independent variants, thus resulting in a total of 167 loci (Supplementary Table 2).

We advanced these loci for replication in a GWAS of refractive error carried out by the UK Biobank Eye & Vision (UKEV) Consortium (n = 95,505)\(^18\) (Methods). Six out of the 167 variants were not considered for replication analysis. One of these five variants (rs3138141, RDHS) was identified previously and therefore still considered a refractive-error risk variant\(^19,20\).

The remaining 161 genetic variants were tested for replication. Among the candidate variants, 86% (138/161) replicated: 104 (65%) replicated surpassing genome-wide significance, and 34 replicated surpassing Bonferroni correction (P < 3.0 × 10\(^{-4}\); 21.1%); another 12 showed nominal evidence for replication (0.05 < P < 3.0 × 10\(^{-4}\); 7.5%); and only 11 (7%) did not replicate at all (Table 1 and Supplementary Table 2).

Because CREAM and 23andMe used different phenotypic outcomes, we evaluated the consistency of genotypic effects by comparing marker-wise additive genetic effect sizes (in diopters per risk-allele variant) for SphE from CREAM-EUR against those (in log(hazard ratio(HR)) per risk-allele variant) for AODM from 23andMe. All variants that were strongly associated with either outcome (P < 0.001) were concordant in direction of effect and had highly correlated effect sizes (Fig. 2a,b and Supplementary Fig. 8). For these variants, a 10% decrease in log(HR) for AODM, indicating an earlier age at myopia onset, was associated with a decrease of 0.15 diopters in SphE. A quantitative analysis of all common SNPs (minor allele frequency (MAF) >0.01; HapMap3) through LD-score regression yielded a genetic correlation of 0.93 (95% confidence interval (CI) 0.86–0.99; P = 2.1 × 10\(^{-9}\)), thus confirming that the effect sizes for both phenotypic outcomes were closely related.

**Gene annotation of susceptibility loci.** We annotated all genetic variants with wANNOVAR by using the University of California Santa Cruz (UCSC) Known Gene database (see URLs\(^21\)). The 139 identified genetic loci were annotated to 208 genes and known transcribed RNA genes (Table 1, Supplementary Table 2 and Methods). The physical positions of the lead genetic variants relative to protein-coding genes are shown in Fig. 1c. 86% of the identified variants were either intragenic or less than 50 kb from the 5’ or 3’ end of the transcription start site. We found seven exonic variants (Supplementary Table 7), of which two had MAF ≤0.05: rs5442 (GNB3) and rs17400325 (PDE11A). The index SNP in the GNB3 locus with MAF 0.05 in Europeans is a highly conserved missense variant (p.Gly272Ser) predicted to be damaging by PolyPhen-2 (ref. 35) and sift\(^36\). PDE11A is presumed to play a role in tumorigenesis, brain function and inflammation\(^37\). The index SNP in the PDE11A locus with MAF 0.03 in Europeans is also a highly conserved missense variant (p.Gly72Ser) predicted to be damaging by PolyPhen-2 (ref. 35) and sift\(^36\). PDE11A is presumed to play a role in tumorigenesis, brain function and inflammation\(^37\). The index SNP in the PDE11A locus with MAF 0.03 in Europeans is also a highly conserved missense variant (p.Tyr727Cys); this variant was predicted to be damaging by PolyPhen-2, sift\(^36\) and align GYVG\(^38,39\). The other exonic variants, rs1064583 (COL10A1), rs807037 (KAZALD1), rs1550094 (PRSS56), rs35337422 (RD3L) and rs6420484 (TSPAN10), were not predicted to be damaging.

The most significant variant (stage 3; rs12193446, P = 4.21 × 10\(^{-48}\)) resides on chromosome 6 within a noncoding-RNA sequence, BC035400, in an intron of the LAMA2 gene. This locus had been identified previously, but our current fine mapping redefined the most associated variant. The function and potential downstream target sites of BC035400 are currently unknown. The previously most strongly associated variant, rs524952 on chromosome 15 near GJD2, was the second most significant variant (P = 2.28 × 10\(^{-46}\)).
Post-GWAS analyses. We performed two gene-based tests, fastBAT and EUGENE, and applied a functional enrichment approach with fgwas (Methods). With fastBAT, we identified 13 genes at $P < 2.0 \times 10^{-6}$, one of which (CHD7) had been identified previously. Using EUGENE, we found seven genes at $P < 2.0 \times 10^{-6}$ after incorporation of blood expression quantitative trait loci (eQTLs). With fgwas, we identified six loci, which were annotated to nine genes, at a posterior probability $>0.9$. Two genes (HMGN4 and TLX1) showed significant associations in two or more approaches. Together, these post-GWAS approaches resulted in a total of 22 additional candidate loci for refractive error, annotated to 25 genes (Supplementary Table 8). These results increase the overall number of significant genetic associations to 161 candidate loci.

Polygenic risk scores. We calculated polygenic risk scores (PGRS) per individual at various $P$ thresholds (Methods) for RSI–III ($n = 10,792$) after recalculating $P$ and $Z$ scores of variants from stage 3 excluding RSI–III. The highest fraction of phenotypic variance (7.8%) was explained with 7,307 variants at a $P$-value threshold of 0.005 (Supplementary Table 9). A PGRS based on these variants distinguished between individuals with hyperopia and myopia at the lower and higher deciles (Fig. 3); those in the highest decile had a 40-fold-greater risk of myopia. When the PGRS was stratified for the median age (<63 or $>63$ years), we found a significant difference in the variance explained (<63 years, 8.9%; $>63$ years, 7.4%; $P = 0.0038$). The variance explained by PGRS was not significantly different between males and females.
Table 1  Results of the meta-analysis of CREAM and 23andMe for the previously identified loci and a subset of the newly identified loci, and replication in UK Biobank

| SNP           | Chromosome | Position | Nearest bid and effect allele | Replication of the HapMap II index variants for refractive error per locus in the stage 3 meta-analysis | Haplotype | Sample size | P | Z-score | Direction | Pvalue | Sample size | P | Z-score | Direction | Pvalue | Sample size | P | Z-score | Direction | Pvalue |
|---------------|------------|----------|------------------------------|--------------------------------------------------------------------------------------------------|-----------|-------------|----|---------|------------|---------|-------------|----|---------|------------|---------|-------------|----|---------|------------|---------|-------------|----|---------|------------|---------|
| rs2155413     | 11         | 84634790 | A/C                          | 0.482/0.655                                           | 0.482     | 2.99        | 10–4  | 10–15   | 159,504   | 0.081   | 160,151     | 8.20 | 10–107   | 0.002    | 160,150 | 3.70        | 10–12   |
| rs4793501     | 17         | 68718734 | T/C                          | 0.575/0.444                                           | 0.575     | 0.592       | 10–12 | 10–18   | 160,150   | 0.02     | 160,147     | 6.00 | 10–10     | 0.003    | 159,504 | 6.70        | 10–10   |
| rs7042950     | 9          | 77149837 | A/G                          | 0.732/0.392                                           | 0.732     | 0.912       | 10–12 | 10–18   | 160,153   | 0.02     | 160,153     | 6.30 | 10–12    | 0.003    | 159,504 | 3.70        | 10–10   |
### Table 1: Results of the meta-analysis of CREAM and 23andMe for the previously identified loci and a subset of the newly identified loci, and replication in UK Biobank (Continued)

| SNP | Chromosome Position | Nearest loci and Effect allele | Other allele(s) | Replication of the HapMap II index variants for refractive error per locus in the stage 3 meta-analysis | Sample-size (n) | East Asian (EA) EAF | EUR (Europe) EAF | Sample-size (n) | EA P-value | EUR P-value | replication |
|-----|---------------------|-------------------------------|----------------|---------------------------------------------------|----------------|-------------------|----------------|----------------|-------------|-------------|------------|
| rs4687586 | 3 | 53087971 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs7534262 | 14 | 60,850,703 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs87322981 | 10 | 8017253 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs9545535 | 5 | 10,017,564 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs9095074 | 10 | 4,945,538 | T | C | 0.447 | 0.568 | 0.782 | 52 | 0.05 | 21,368 | 0.568 |
| rs13194824 | 16 | 65,290,364 | C | G | 0.684 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs13061904 | 14 | 1,293,365 | C | G | 0.684 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs15568647 | 11 | 3,082,924 | T | C | 0.328 | 0.649 | 0.568 | 52 | 0.05 | 2,218 | 0.568 |
| rs13062035 | 12 | 7,382,658 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs11101231 | 11 | 17,116,880 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs18258091 | 5 | 18,362,865 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs7747 | 4 | 8,082,702 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs4687586 | 3 | 53087971 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs7534262 | 14 | 60,850,703 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs87322981 | 10 | 8017253 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs9545535 | 5 | 10,017,564 | C | G | 0.605 | 5.86×10^-10 | 6.64×10^-10 | 739 | 0.05 | 150,277 | 160×10^-10 |
| rs9095074 | 10 | 4,945,538 | T | C | 0.447 | 0.568 | 0.782 | 52 | 0.05 | 21,368 | 0.568 |
| rs13194824 | 16 | 65,290,364 | C | G | 0.684 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs13061904 | 14 | 1,293,365 | C | G | 0.684 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs15568647 | 11 | 3,082,924 | T | C | 0.328 | 0.649 | 0.568 | 52 | 0.05 | 2,218 | 0.568 |
| rs13062035 | 12 | 7,382,658 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs11101231 | 11 | 17,116,880 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs18258091 | 5 | 18,362,865 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |
| rs7747 | 4 | 8,082,702 | C | G | 0.695 | 0.295 | 0.909 | 40 | 0.05 | 3,191 | 0.568 |

We identified 140 loci for refractive error with genome-wide significance (P < 5×10^-8) on the basis of the meta-analysis of the genomewide-wide association and replication performed in 160,242 participants in three cohorts (CREAM, CREAM ERF, and UK Biobank). The effect size was standardized per standard deviation (SD) of refractive error. For each locus, we identified the effect allele, other alleles, effect allele frequency (EAF), and P-value. We also performed replication in the UK Biobank. Bolded values indicate that the effect size was statistically significant at P < 5×10^-8. SNPs in bold indicate that the effect size was statistically significant at P < 5×10^-8.
Fig. 2 | Correlation of statistical significance and effect size of SNPs on the basis of SphE in diopters and AODM in years. a, P comparison of all genetic variants with $P < 1.0 \times 10^{-3}$ (n = 7,249) between CREAM meta-analysis (stage 1) and 23andMe (stage 2) meta-analysis. Shown is the overlap (green) and the difference (purple) in P signals per cohort for genetic variants. Purple genetic variants are only genome wide significant (GWS) in either CREAM or 23andMe. Black, genetic variants with $P$ between $5.0 \times 10^{-3}$ and $1.0 \times 10^{-3}$ in both CREAM and 23andMe. b, Comparison of effects (SphE and logHR of AODM in years; $P < 1.0 \times 10^{-3}$, n = 7,249) between CREAM and 23andMe, with color code as in a. The effects were concordant in their direction of effect on refractive error. We performed a simple linear regression between the effects of CREAM and 23andMe; the regression slope was -0.15 diopters per logHR of AODM in years.

Fig. 3 | Risk of refractive error per decile of polygenic risk score (Rotterdam Study I–III, n = 10,792). Distribution of refractive error in subjects from RSI–III (n = 10,792) as a function of the optimal polygenic risk score (including 7,303 variants at $P \leq 0.005$ explaining 7.8% of the variance of SphE; Supplementary Table 9). The mean odds ratio (OR) of myopia (black line) was calculated per polygenic-risk-score category by using the lowest category as a reference. High myopia, SphE $> 6$ diopters; moderate myopia, SphE $> 6$ diopters and $\leq 3$ diopters; low myopia, SphE $< 1.5$ diopters; and $< 1.5$ diopters; emmetropia, SphE $\geq 1.5$ diopters and $\leq 1.5$ diopters; low hyperopia, SphE $> 1.5$ diopters and $< 3$ diopters; moderate hyperopia, SphE $\geq 3$ diopters and 6 diopters; high hyperopia, SphE $\geq 6$ diopters. (R), reference group.

(8.3 vs. 7.5%, respectively; $P = 0.13$). The predictive value (area under the receiver operating characteristic curve) of the PGRS for myopia vs. hyperopia, adjusted for age and sex, was 0.77 (95% CI = 0.75–0.79), a value 10% higher than previous estimations.

Transcultural comparison of genotypic effects. To explore potential ancestry differences in the identified refractive-error loci, we calculated the heritability explained by common genetic variants (SNP-$h^2$) for Europeans and Asians, by using LD-score regression. The SNP-$h^2$ was 0.214 (95% CI 0.185–0.243) and 0.172 (95% CI 0.154–0.190) in the European samples (CREAM-EUR and 23andMe, respectively), but was only 0.053 (95% CI 0.025–0.131) in the Asian sample (CREAM-EAS). Next, we estimated the genetic correlation between Europeans and Asians by comparing variant effect sizes for common variants in Popcorn (Methods). Two genetic correlation metrics were calculated: (i) a genetic-effect correlation ($\rho_{ge}$) that quantifies the correlation in SNP effect sizes between Europeans and Asians without taking into account ancestry-related differences in allele frequency and (ii) a genetic-impact correlation ($\rho_{gi}$) that estimated the correlation in variance-normalized SNP effect sizes between the two ancestry groups (Table 2). Estimates of $\rho_{ge}$ were high between Europeans and Asians, but were significantly different from 1 (0.79 and 0.80, respectively, at $P < 1.9 \times 10^{-4}$; Table 2), thus indicating a clear genetic overlap but a difference in per-allele effect size. Estimates of $\rho_{gi}$ were similarly high (>0.8) but were not significantly different from 1 for the correlation between CREAM-EUR and CREAM-ASN ($P = 0.065$), thus indicating that the genetic impact of these alleles may still be similar.

In silico pathway analysis. We used an array of bioinformatics tools to investigate potential functions and pathways of the associated genes. We first used DEPICT to perform a gene set enrichment analysis, a tissue-type enrichment analysis and a gene prioritization analysis, on all variants with $P < 1.0 \times 10^{-5}$ from stage 3. The gene set enrichment analysis resulted in 66 reconstituted gene sets, of which 55 (83%) were eye related. To decrease redundancy among pathways, we clustered the significant pathways into 13 meta-gene sets (false discovery rate (FDR) <5% and $P < 0.05$) (Supplementary Note, Fig. 4 and Supplementary Table 10). The most significant gene set was ‘abnormal photoreceptor inner segment morphology’ (Mammalian Phenotype Ontology (MP) 0003730; $P = 1.79 \times 10^{-10}$). The eye-related meta-gene sets consisted of ‘thin retinal outer nuclear layer’ (MP 0008515; 27 (55%) gene sets), ‘detection of light stimulus’ (Gene Ontology (GO) 0009583; 13 (24%) gene sets), ‘nonmotile primary cilium’ (GO 0031513; 4 (6%) gene sets) and ‘abnormal
The first three meta-gene sets had a Pearson’s correlation ≥0.6. Interestingly, RGR, RP1LI, RORB and GNB3 were present in all of these meta-gene sets. The retina was the most significant tissue of expression according to the tissue-type enrichment analysis (P = 1.11 × 10^{-4}, FDR < 0.01). From the gene prioritization according to the genetic impact (P value ρgi) of the meta-gene set ‘detection of light stimulus’ (GO 0009583).
equally rated categories (Methods, Fig. 5, Supplementary Table 11) genes with biologically plausible roles in eye growth. We used ten one point could be assigned for ‘annotation’, even though it has four columns (i.e., a genetic variant is located in only one of these four categories).

pathway analysis). We assessed genes bearing drug targets (salmon red) but did not assign a scoring point to that category. Asterisk indicates that only phenotype in mice, ocular phenotype in humans); pathways (green; DEPICT gene set enrichment, DEPICT gene-prioritization analysis and IPA canonical an RNA structure); expression (yellow; eQTL, expression in adult human ocular tissue, expression in developing ocular tissue); biology (dark yellow; ocular genetic variant bearing an exonic protein-altering variant or non-protein-altering variant, genetic variant residing in a 5′ or 3′ UTR of a gene or transcribing an RNA structure); expression (yellow; eQTL, expression in adult human ocular tissue, expression in developing ocular tissue); biology (dark yellow; ocular phenotype in mice, ocular phenotype in humans); pathways (green; DEPICT gene set enrichment, DEPICT gene-prioritization analysis and IPA canonical pathway analysis). We assessed genes bearing drug targets (salmon red) but did not assign a scoring point to that category. Asterisk indicates that only one point could be assigned for ‘annotation’, even though it has four columns (i.e., a genetic variant is located in only one of these four categories).

![Table 1: Gene Priority Score](image)

| Locus     | Locus name       | Gene priority score | Internal replication (≥2 cohorts) | Exonic protein altering | Exonic non–protein altering | RNA (inc, ins, inc/ex) | eQTL | Expression in human adult ocular tissue | Expression in human developing ocular tissue (fetal–24 weeks) | Ocular phenotype in mice | Ocular phenotype in humans | DEPICT gene set enrichment | DEPICT gene prioritization | IPA canonical pathways |
|-----------|------------------|---------------------|-----------------------------------|-------------------------|-----------------------------|-------------------------|-------|------------------------------------------|-------------------------------------------------------------|------------------------|----------------------------|---------------------------|--------------------------|-------------------------|
| GNB3      | GNB3             | 8                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| RDH5      | BLOC1S1–RDH5, RDH5 | 7                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| CYP26A1   | CYP26A1, MYOF    | 7                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| ROR8      | ROR8             | 7                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| RGR       | RGR              | 7                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| TJP2      | TJP2             | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| PRSS56    | PRSS56           | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| CABP4     | CABP4            | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| FBN1      | FBN1             | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| GJD2      | GJD2, GOLGA8B    | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| KCN2      | KCN2             | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| KCNMA1    | KCNMA1           | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| MAF       | DYNLRB2, MAF     | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| RCBTB1    | RCBTB1           | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| ST18      | FAM150A, ST18    | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| TCF7L2    | TCF7L2           | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |
| ZEB2      | ZEB2             | 6                   |                                   |                         |                             |                         |       |                                          |                                                             |                        |                            |                           |                          |                        |

Fig. 5 | Genes ranked according to biological and statistical evidence. Genes ranked (orange) according to ten equal categories that can be grouped into the following: internal replication of genetic variant in two or more cohorts (purple; CREAM-EUR, CREAM-ASN and/or 23andMe); annotation (light blue; genetic variant bearing an exonic protein-altering variant or non-protein-altering variant, genetic variant residing in a 5′ or 3′ UTR of a gene or transcribing an RNA structure); expression (yellow; eQTL, expression in adult human ocular tissue, expression in developing ocular tissue); biology (dark yellow; ocular phenotype in mice, ocular phenotype in humans); pathways (green; DEPICT gene set enrichment, DEPICT gene-prioritization analysis and IPA canonical pathway analysis). We assessed genes bearing drug targets (salmon red) but did not assign a scoring point to that category. Asterisk indicates that only one point could be assigned for ‘annotation’, even though it has four columns (i.e., a genetic variant is located in only one of these four categories).

to DEPICT, seven genes were highlighted as the most likely causal genes at P < 7.62 × 10−8 and FDR < 0.05: ANO2, RP1L1, GNB3, EDN2, RORB and CABP4.

Next, we performed a canonical pathway analysis on all genes annotated to the variants of stage 3, by using Ingenuity Pathway Analysis (IPA; see URLs). All genes were run against the IPA database incorporating functional biological evidence on genomic and proteomic expression according to regulation or binding studies. IPA identified ‘glutamate receptor signaling’ with the central player NF-κB as the most significant pathway after correction for multiple testing (ratio of the number of molecules, 8.8%; Fisher’s exact P = 1.56 × 10−4; Supplementary Fig. 9).

From disease-associated loci to biological mechanisms. We adapted the scoring scheme designed by Fritsche et al.49 to highlight genes with biologically plausible roles in eye growth. We used ten equally rated categories (Methods, Fig. 5, Supplementary Table 11 and Supplementary Note). We found that 109 index variants replicated in two or more individual cohorts; there was evidence for seven genetic variants with eQTL effects in multiple tissue types; nine exonic variants, seven of which predicted protein alterations (Supplementary Table 7); 31 RNA genes, five of which were located in the 3′ or 5′ untranslated region (UTR) (Supplementary Table 12 and Supplementary Fig. 10); 84 genes resulting in an ocular phenotype in humans (Supplementary Table 13) and 36 in mice (Supplementary Table 14); 172/212 (81%) genes expressed in human ocular tissue (Supplementary Note and Supplementary Table 15); 41 genes identified by DEPICT at P < 5.4 × 10−4 and FDR < 0.05; and 45 genes that contributed to the most significant canonical IPA pathways. Notably, 48 of the associated genes encode known drug targets (Supplementary Table 16).

The gene with the highest biological-plausibility score (score = 8) was GNB3, a highly conserved gene encoding a G-protein-coupled receptor expressed in rod and cone photoreceptors and ON bipolar cells. GNB3 participates in signal transduction through G-protein-coupled receptors and enhances the temporal accuracy
of phototransduction and ON-center signaling in the retina\(^a\). As described above, the index SNP contains a missense variant associated with refractive errors. Nonsynonymous mutations within GNB3 are known to cause syndromic congenital stationary night blindness\(^b\) in humans; progressive retinopathy and globe enlargement in chickens\(^c\); and abnormal development of the photoreceptor-bipolar synapse in knockout mice\(^d\).\(^e\).

Other highly ranked (score of 7) genes included CYP26A1, GRIA4, RDH5, RORB and GNB3, all previously associated with refractive error, and one newly identified gene, EFEMP1. EFEMP1 encodes a member of the fibrillin family of extracellular-matrix glycoproteins and is found panocularly, including in the inner nuclear layer and Bruch's membrane. Mutations in this gene lead to specific macular dystrophies\(^f\), whereas variants have also been shown to cosegregate with primary open-angle glaucoma\(^g\) and to be associated with optic disc cup area\(^h\).

Several other genes were noteworthy for their function. CABP4, which encodes a calcium-binding protein expressed in cone and rod photoreceptor cells, mediates Ca\(^{2+}\) influx and glutamate release in the photoreceptor bipolar synapse\(^i\). Mutations in this gene have been described in congenital cone-rod synaptic disorder\(^j\), a retinal dystrophy associated with nystagmus, photophobia and high hyperopia. KCNMA1 encodes pore-forming alpha subunits of Ca\(^{2+}\)–activated K\(^+\) channels. These channels regulate synaptic transmission exclusively in the rod pathway\(^k\). ANO2 encodes a Ca\(^{2+}\)–activated Cl\(^-\) channel recently reported to regulate retinal pigment epithelium (RPE) cell volume in a light-dependent manner\(^l\). EDN2 encodes a potent vasoconstrictor that binds to two G-protein-coupled receptors encoded by EDNRA, which resides on bipolar dendrites, and the protein product of EDNRB, which is present on Mueller and horizontal cells. Both receptors are also present on chorioidal vessels\(^m\), thus implying that the choroid as well as retinal cells are target sites of this gene. RPL21 is expressed in cone and rod photoreceptors, where it is involved in the maintenance of microtubules in the connecting cilium\(^n\). Mutations in this gene cause dominant macular dystrophy and retinitis pigmentosa\(^o\). We replicated two genes involved in myopia in family studies: (i) FBN1, which bears mutations causing Marfan (MIM 154700) and Weil Marchesani (MIM 608328) syndromes, and (ii) PTPRR, one of the candidates in the MYP3 locus, which was identified on the basis of linkage in families with high myopia\(^p\).

The location of rs7449443 (\(P = 3.58 \times 10^{-8}\)) is notable because it resides between DRD1 and LINC01951. DRD1 encodes dopamine receptor 1 and is known to modulate dopamine receptor 2–mediated events\(^q\). The dopamine pathway has been implicated in myopia pathogenesis in many studies\(^r\).\(^s\),\(^t\). SNPs in and near other genes involved in the dopamine pathway (dopamine receptor binding, synthesis, degradation and transport)\(^u\),\(^v\),\(^w\) did not show genome-wide-significant associations (Supplementary Note, Supplementary Table 17 and Supplementary Fig. 11).

There were 31 genetic variants in or near DNA structures transcribing RNA genes (noncoding RNA, long intergenic noncoding RNAs, tRNAs, small nucleolar RNAs and ribosomal RNAs). Notably, five were in the transcription region, and 13 were in the vicinity (\(>0\) kb and \(\leq50\) kb) of the start or end of the transcription region. They received low scores because many have no reported function or disease association to date (Fig. 5, Supplementary Fig. 10).
and Supplementary Table 12). Our ranking of genes according to functional information existing in the public domain does not necessarily represent the true order of importance for refractive-error pathogenesis. The observation that genes with strong statistical association were distributed over all scores supports this concept. Nevertheless, this list may aid in selection of genes for subsequent functional studies.

Finally, integration of all our findings together with literature allowed us to annotate a large number of genes to ocular cell types (Fig. 6). All cell types of the retina contained refractive-error genes, as well as RPE, vascular endothelium and extracellular matrix.

Genetic pleiotropy. We performed a GWAS catalog lookup, using FUMA to investigate the overlap of genes with other common traits (Supplementary Fig. 12). Refractive error and hyperopia were replicated significantly after correction for multiple testing (adjusted \( P \text{-value} = 1.44 \times 10^{-52} \) and \( 9.34 \times 10^{-41} \), respectively). We found significant overlap with 74 other traits, of which height (adjusted \( P \text{-value} = 1.11 \times 10^{-10} \)), obesity (adjusted \( P = 1.38 \times 10^{-10} \)) and body mass index (adjusted \( P = 4.05 \times 10^{-7} \)) were most important. Ocular diseases significantly associated were glaucoma (optic cup area and intraocular pressure, adjusted \( P = 2.69 \times 10^{-9} \) and \( 3.01 \times 10^{-5} \), respectively) and age-related macular degeneration (adjusted \( P = 1.27 \times 10^{-5} \)).

Discussion

Myopia may become the leading cause of blindness worldwide in the near future, which suggests a grim outlook for which current counteractions remain insufficient\(^{11,12}\). To improve understanding of the genetic landscape and biology of refractive error, we conducted a large GWAS meta-analysis in 160,420 participants of mixed ancestry with replication in 95,505 participants. This study led to the identification of 139 independent susceptibility loci through single-variant analysis and 22 additional loci through post-GWAS methods, representing a fourfold increase in refractive-error genes. Most annotated genes were found to be expressed in the human posterior segment of the eye. Using in silico analysis, we identified significant biological pathways, of which retinal cell physiology, light processing and, specifically, glutamate receptor signaling were the most prominent mechanisms. Our integrated bioinformatic approach highlighted known ocular functionality for many genes.

To ensure the robustness of our genetic associations, we included studies of various designs and populations; sought replication in an independent cohort of significant sample size; and stringently accounted for population stratification by performing genomic control at all stages of the meta-analysis\(^{11,12}\). We combined studies with outcomes based on actual refractive-error measurements, as well as on the self-reported age of myopia onset, and found the direction of effect of the associated variants, as well as their effect size, to be highly consistent. Combining two different outcome measures may appear unconventional, but age of onset and refractive error have been shown to be very tightly correlated\(^{11,24,74,75}\). Moreover, the high genetic correlation (93%) of common SNPs between the two phenotypes underscores their similarity. The most compelling evidence was provided by replication of 86% of the discovered variants in the independent UKEV data, which also used conventional refractive-error measurements. This robustness indicates that both phenotypic outcomes can be used to capture a shared source of genetic variation. In addition, we found transancestral replication of significant loci and a high correlation of genetic effects of common variants in Europeans and Asians. Our findings support a largely shared genetic predisposition to refractive error and myopia in the two ancestries, although ancestry-specific allelic effects may exist. The low heritability estimate in Asians may be partly explained by the low representation of this ancestral group in our study sample; alternatively, it may imply that environmental factors explain a greater proportion of the phenotypic risk and recent rise in myopia prevalence in this ancestry group\(^{16}\).

Limitations of our study were the possibility of false-negative findings due to genomic control and underrepresentation of studies including individuals of Asian ancestry. The heterogeneity of the observed effect estimates was large for several associated variants, but this result was not unexpected, given the large number of collaborating studies with varying methodology.

Although neurotransmission was a previously suggested pathway\(^{26,27}\), our current pathway analyses provide more in-depth insights into the retinal circuitry driving refractive error. DEPICT identified ‘thin retinal outer nuclear layer’, ‘detection of light stimulus’ and ‘nonmotile primary cilium’ as the most important meta-gene sets. These are the main characteristics of photoreceptors, which are located in the outer retina and contain cilia. These photosensitive cells drive the phototransduction cascade in response to light, which in turn induces visual information processing. IPA indicated ‘glutamate receptor signaling’ as the most significant pathway. Glutamate is released by photoreceptors and determines conductance of retinal signaling to the ON and OFF bipolar cells\(^{28}\). Our functional gene lookups provide evidence that rod (CLU) as well as cone (GNB3) bipolar cells play a role. Together, these findings strongly suggest that light response and light processing in the retina are initiating factors leading to refractive error.

The genetic association with light-dependent pathways may also be linked to the well-established protective effect of outdoor exposure on myopia. We found evidence suggesting a genetic association with DRD1. The dopaminergic pathway has been studied extensively in animal models for its role in controlling eye growth in response to light\(^{89,90,91,92}\). DRD1 has been found to be a mediator in this process, because bright light increases DRD1 activity in the bipolar ON pathway, and diminishes form-deprivation myopia in mice. Blockage of DRD1 reverses this inhibitory effect\(^{89}\). We did not find evidence of direct involvement of other genes in the dopamine pathway, but GNB3 may be an indirect modifier, because it is a molecule involved in dopamine downstream signaling and has been shown to influence the availability of the dopamine transporter DAT\(^{89}\). Although it is a promising target for therapy, further evidence of DRD1 in human myopia genesis is warranted.

Novel pathways implicated by the newly identified genes are anterior-segment morphology (TCF7L2, VIPR2 and MAF) and angiogenesis (FLT1). In addition, the high number of variants residing near genes encoding small RNAs suggests that post-transcriptional regulation is an important mechanism, because these RNAs are known to play a distinct and central regulatory role in cells\(^{93}\). These findings should serve as leads for future studies performing detailed mapping of cellular networks as well as for functional studies on genes that have been implicated in ocular phenotypes, that have protein-altering variants and that are proven drug targets.

Our evaluation of shared genetics between refractive error and other disease-relevant phenotypes highlighted overlap with anthropometric traits such as height, obesity and body mass index. These findings may provide valuable additional clues regarding the phenotypic outcomes of perturbations of some of the networks identified.

Our genetic observations add credence to the current notion that refractive errors are caused by a retina-to-sclera signaling cascade that induces scleral remodeling in response to light stimuli. The concept of this cascade originates from various animal models showing that form deprivation, retinal defocus and contrast, ambient light and wavelength influence eye growth in young animals\(^{94,95}\). The cell-specific moieties in this putative signaling cascade in humans are largely unknown, although animal models have implicated GABA, dopamine, all-trans retinoic acid and TGF-β (refs 65,67,74,96). Our study provides a large number of new molecular candidates for this...
cascade and clearly implicates a wide range of neuronal cell types in the retina, the RPE, the vascular endothelium and components of the extracellular matrix. The many interprotein relationships exemplify the complexity of eye growth and provide a challenge to developing strategies to prevent pathological eye elongation.

In conclusion, by using a cross-ancestry design in a large study population on common refractive errors, we identified numerous novel loci and pathways involved in eye growth. Our multidisciplinary approach incorporating GWAS data with in silico analyses and expression experiments provides an example for the design of future genetic studies for complex traits. Additional genetic insights into refractive errors will be gained by increasing sample size and genotyping depth; by performing family studies to identify rare alleles with large effects; and by evaluating population extremes. Our list of plausible genes and pathways provides a plethora of data for future studies focusing on gene–environment interaction and on translation of GWAS findings into starting points for therapy.

URLs: LDSC, https://github.com/bulik/lodsc/; Popcorn, https://github.com/brielin/Popcorn/; Online Mendelian Inheritance in Man (OMIM), http://omim.org/; wANNOVAR, http://wannovar.wgbli.org/; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/; SIFT, http://sift.jcvi.org/www/SIFT_aligned_seqs_submit.html; MutationTaster, http://www.mutationtaster.org/; IPA, https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/; 1000 Genomes Project (release 2 May 2013), ftp.1000genomes.ebi.ac.uk/UCSC Genome Browser, https://genome.ucsc.edu/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0127-7.

Received: 22 May 2017; Accepted: 26 March 2018; Published online: 28 May 2018

References

1. Pan, C. W., Ramamurthy, D. & Saw, S. M. Worldwide prevalence and risk factors for myopia. *Optom. Physiol. Opt.* 32, 3–16 (2012).
2. Morgan, I. G. What public policies should be developed to deal with the epidemic of myopia? Option. Vis. Sci. 93, 1058–1060 (2016).
3. Morgan, I. G. & Rose, K. How genetic is school myopia? *Prog. Retin. Eye Res.* 24, 1–38 (2005).
4. Morgan, I. G., Ohno-Matsui, K. & Saw, S. M. Myopia. *Lancet* 379, 1739–1748 (2012).
5. Williams, K. M. et al. Increasing prevalence of myopia in Europe and the impact of education. *Ophthalmology* 122, 1489–1497 (2015).
6. Williams, K. M. et al. Prevalence of refractive error in Europe: the European Eye Epidemiology (E3) Consortium. *Eur. J. Epidemiol.* 30, 305–315 (2015).
7. Vongphanit, J., Mitchell, P. & Wang, J. I. Prevalence and progression of myopic retinopathy in an older population. *Ophthalmology* 109, 704–711 (2002).
8. Seet, B. et al. Myopia in Singapore: taking a public health approach. *Br. J. Ophthalmol.* 85, 521–526 (2001).
9. Smith, T. S., Frick, K. D., Holden, B. A., Fricke, T. R. & Naidoo, K. S. Potential lost productivity resulting from the global burden of uncorrected refractive error. *Bull. World Health Organ.* 87, 431–437 (2009).
10. Verhoeven, V. J. et al. Visual consequences of refractive errors in the general population. *Ophthalmology* 122, 101–109 (2015).
11. Tideman, J. W. et al. Association of axial length with risk of uncorrectable visual impairment for Europeans with myopia. *JAMA Ophthalmol.* 134, 1355–1363 (2016).
12. Flitcroft, D. I. The complex interactions of retinal, optical and environmental factors in myopia aetiology. *Prog. Retin. Eye Res.* 31, 622–660 (2012).
13. Nakanshi, H. et al. A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1. *PLoS Genet.* 5, e1000660 (2009).
14. Lam, C. Y. et al. A genome-wide scan maps a novel high myopia locus to 5p15. *Invest. Ophthalmol. Vis. Sci.* 49, 3768–3778 (2008).
15. Stambolian, D. et al. Meta-analysis of genome-wide association studies in five cohorts reveals common variants in RBFOX1, a regulator of tissue-specific splicing, associated with refractive error. *Hum. Mol. Genet.* 22, 2754–2764 (2013).
16. Fan, Q. et al. Genetic variants on chromosome 1q41 influence ocular axial length and high myopia. *PLoS Genet.* 8, e1002753 (2012).
17. Fan, Q. et al. Meta-analysis of gene-environment-wide association scans accounting for education level identifies additional loci for refractive error. *Nat. Commun.* 7, 11008 (2016).
18. Cheng, C. V. et al. Nine loci for ocular axial length identified through genome-wide association studies, including shared loci with refractive error. *Am. J. Hum. Genet.* 93, 264–273 (2013).
19. Shi, Y. et al. Exome sequencing identifies ZNF644 mutations in high myopia. *PLoS Genet.* 7, e1002084 (2011).
20. Shi, Y. et al. Genetic variants at 13q12.12 are associated with high myopia in the Han Chinese population. *Am. J. Hum. Genet.* 88, 805–813 (2011).
21. Li, Y. J. et al. Genome-wide association studies reveal genetic variants in CTNND2 for high myopia in Singapore Chinese. *Ophthalmology* 118, 368–375 (2011).
22. Li, Z. et al. A genome-wide association study reveals association between common variants in an intergenic region of 4p25 and high-grade myopia in the Chinese Han population. *Hum. Mol. Genet.* 20, 2861–2868 (2011).
23. Liu, J. & Zhang, H. X. Polymorphism in the 11q24.1 genomic region is associated with myopia: a comprehensive genetic study in Chinese and Japanese populations. *Mol. Vis.* 20, 352–358 (2014).
24. Sun-Vieto, N. et al. Mutations in SOCO2 are associated with autosomal-dominant high-grade myopia. *Am. J. Hum. Genet.* 92, 820–826 (2013).
25. Aldaehme, M. A. et al. Mutations in LRMP1 are associated with severe myopia in humans. *Am. J. Hum. Genet.* 93, 313–320 (2013).
26. Verhoeven, V. J. et al. Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nature* 542, 291–295 (2015).
27. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 19, 807–812 (2011).
28. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826 (2017).
29. Plotnikov, D., Guggenheim, J. & The UK Biobank Eye and Vision Consortium. Is a large eye size a risk factor for myopia? A Mendelian randomization study. https://www.biobrix.org/content/early/2017/12/29/240283/ (2017).
30. Hsu, F. et al. The UCSC Known Genes. *Bioinformatics* 22, 1036–1046 (2006).
31. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249 (2010).
32. Noll, C. F. & Henikoff, S. SIFT predicts amino acid changes that affect protein function. *Nucleic Acids Res.* 38, 3812–3814 (2003).
33. Kelly, M. P. Does phosphodiesterase 11A (PDE11A) hold promise as a future therapeutic target? *Curr. Pharm. Des.* 21, 389–416 (2015).
34. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081 (2009).
35. Mathe, E. et al. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* 34, 1317–1325 (2006).
36. Tavtigian, S. V. et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* 43, 295–305 (2006).
37. Bakshi, A. et al. Fast set-based association analysis using summary data from GWAS identifies novel loci for human complex traits. *Sci. Rep.* 6, 32894 (2016).
38. Ferreira, M. A. et al. Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. *J. Allergy Clin. Immunol.* 139, 1148–1157 (2017).
39. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* 94, 559–573 (2014).
40. Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752 (2009).
41. Verhoeven, V. J. et al. Large scale international replication and meta-analysis study confirms association of the 15q14 locus with myopia. The CREAM consortium. *Hum. Genet.* 131, 1467–1480 (2012).
46. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).
47. Brown, C. R. Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. Am. J. Hum. Genet. 99, 76–88 (2016).
48. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. 6, 5890 (2015).
49. 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 (2016).
50. Ritchie, E. R. et al. Vision-guided ocular growth in a mutant chicken model with diminished visual acuity. Exp. Eye Res. 102, 59–69 (2012).
51. Vincent, A. et al. Biallelic mutations in GNBP3 cause a unique form of autosomal-recessive congenital stationary night blindness. Am. J. Hum. Genet. 98, 1011–1019 (2011).
52. Blake, J. A. et al. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. 45, D723–D729 (2017).
53. Springelkamp, H. et al. ARHGEF12 influences the risk of glaucoma by increasing intraocular pressure. Hum. Mol. Genet. 24, 2689–2699 (2015).
54. Stone, E. M. et al. A single EFEMP1 mutation associated with both Leventinese and Doyne honeycomb retinal dystrophy. J. Neurosci. 38, 8123–8133 (2018).
55. Mackay, D. S., Bennett, T. M. & Shiels, A. Exome sequencing identifies a β2-adrenoceptor polymorphism associated with diminished visual acuity. Vision Res. 143, 192–200 (1999).
56. Littink, K. W. et al. A novel homozygous nonsense mutation in CABP4 causes congenital cone-rod synaptic disorder. Invest. Ophthalmol. Vis. Sci. 56, 2076–2086 (2015).
57. Grimes, W. N., Li, W., Chávez, A. E. & Diamond, J. S. BK channels mediate pre- and postsynaptic signaling at retinal synapses in retina. Nat. Neurosci. 12, 585–592 (2009).
58. Keckes, S., Reichhart, N., Noutbe, C. & Straub, O. A. Anatomica (TMEM16B) is in the Ca2+–activated Cl− channel in the retinal pigment epithelium. Exp. Eye Res. 154, 139–150 (2017).
59. Prasanna, G., Narayan, S., Krishnamoorthy, R. R. & Yorio, T. Eyeing endotheilins: a cellular perspective. Mol. Cell. Biochem. 253, 71–88 (2003).
60. Yamashita, T. et al. Essential and synergistic roles of RP1 and RP1L1 in rod photoreceptor synaptic function. Nat. Neurosci. 7, 1079–1087 (2004).
61. Littink, K. et al. A novel homozygous nonsense mutation in CAPP4 causes congenital cone-rod synaptic disorder. Invest. Ophthalmol. Vis. Sci. 50, 2344–2350 (2009).
62. Ashby, R., McCarthy, C. S., Maleszka, R., Megaw, P. & Morgan, I. G. A muscarinic cholinergic antagonist and a dopamine agonist rapidly increase ZENK mRNA expression in the form-deprived chicken retina. Exp. Eye Res. 85, 15–22 (2007).
63. Asano, M. & Schaeffel, F. Animal studies and the mechanism of myopia–protection by light? Optom. Vis. Sci. 93, 1052–1054 (2016).
64. Rymer, J. & Wildsoet, C. F. The role of the retinal pigment epithelium in eye growth regulation and myopia: a review. Vis. Neurosci. 22, 251–265 (2005).
65. Ross, S. et al. Bright light suppresses form-deprivation myopia development with activation of dopamine D1 receptor signaling in the ON pathway in retina. Invest. Ophthalmol. Vis. Sci. 58, 2306–2316 (2017).
66. Chen, P. et al. Effects of CR2ST polymorphism of the GNB3 gene on availability of dopamine transporter in healthy volunteers: a SPECT study. Exp. Eye Res. 123, 16–27 (2015).
67. Scott, M. S. & Ono, M. From snRNA to miRNA: dual function regulatory non-coding RNAs. Biochimie 93, 1987–1992 (2011).
68. McFadden, S. A. Understanding and treating myopia: what more we need to know and future research priorities. Optom. Vis. Sci. 93, 1061–1063 (2016).
69. Smith, E. L. III, Hung, L. F. & Arumugam, B. Visual regulation of refractive development: insights from animal studies. Eye (Lond.) 28, 180–188 (2014).
70. Zhang, Y. & Wildsoet, C. F. RPE and choroid mechanisms underlying ocular growth and myopia. Prog. Mol. Biol. Transl. Sci. 134, 221–240 (2015).
71. Harper, A. R. & Summers, J. A. The dynamic sclera: extracellular matrix remodeling in normal ocular growth and myopia development. Exp. Eye Res. 133, 100–111 (2015).
72. Summers, J. A. The choroid as a sclera growth regulator. Exp. Eye Res. 114, 120–127 (2013).

Acknowledgements
We gratefully thank all study participants, their relatives and the staff at the recruitment centers for their invaluable contributions. We thank all contributors to the CREAM Consortium, 23andMe and UKEye for their generosity in sharing data and help in the production of this publication. Funding for this particular GWAS meta-analysis was provided by the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (grant 648268), the Netherlands Organisation for Scientific Research (NWO, grant 91815655) and the National Eye Institute (grant R01EY020483). Funding agencies that facilitated the execution of the individual studies are acknowledged in the Supplementary Note.

Author contributions
M.S.T., V.J.M.V., S.M., J.A.G., A.I.I., R.W., P.G.H., A.I.I. and E.M.v.L. performed the analyses. C.W.C.K., V.J.M.V., M.S.T., R.W., J.A.G. and S.M. drafted the manuscript, and C.J.H., P.G.H., A.P.K., C.M.v.d.D., D.S., E.M.v.L., J.E.B. W., J.Y.T., N.A.F., Q.S., S-M.S. and V.V. critically reviewed the manuscript. A.-N.K., A.T., C.B., C. Gieger, C.L.S., C.-Y. F., V. Bino, G.C.-P., J.E.B.W., J.E.H., I. S. Ried, J.W., J.W., M.K.W., K.Y., P.M.C., S.M.H., M.S.T., N.A.F., N.E., P.C., P. Gharakhilani, P.K.I., Q.E., R. Höhn, R.L.S., R.P.L., R.W., T.-H.-S.A.-T., T.V., V.W., Y.S., W.Z., X.L.S., Y.C.T., Y.S. and Y.Y.T. performed data analysis for the individual studies; A.D.P., A.G.U., A.T., A.W.H., B.E.K.K., C.C.W.K., C.D., C. Grañal, C.H., C.J.W., C.-Y.C., D.A.M., F.R., G. Benecic, H.-M. H., J.A.G., J.B.I., J.E.B.-W., J.E.G., J.E.W., J.H.L., J.R.V., J.-S. R. Bahi, J. S. Ried, Y.Y.T., K.Y., M.A.M.-S.,
N.G.M., N.P. O. Polo, O. Párisen, O. T.R., P. Gupta, P.J.E., P.M., P.N.B., R.K., S.K.I., S.-M.S., T.L., T.M., W.Z., Y.C.T. and Y.X.W. contributed to data assembly. A.A.B.B., A.W., C. Grazal, D.S., K.N.W., S.W.T. and T.L.Y. performed expression experiments, and M.S.T., A.A.B.B., R.J.-d.S. and R. Hask performed in silico pathway analyses. C.C.W.K. and C.I.H. conceived and designed the outline of the current report, and supervised the conduction of experiments and analyses jointly with A.M., A.W.H., C.D., C.H., C.I.H., C.M.v.d.W., C.W., C.-Y.C., D.A.M., D.S., E.-S.T., E.M., G. Biino, I.R., J.A.G., J.B.I., J.E.B.-W., J.E.C., J.F.W., J.H.I., J.R.V., J.Y.T., N.A., N.A.F., N.P. O. Párisen, O.T.R., P.J.E., P.N.B., S.K.I., S.-M.S., T.L.Y., T.W., T.L.Y., V.V., Y.X.W. and Y.T. M.P.C. analyzed the data and performed statistical analyses. The 23andMe research team, CREAM and the UK Biobank Eye and Vision Consortium contributed reagents/materials/analysis tools and performed statistical analyses.

Competing interests
N.A.F., N.E., J.Y.T. and the 23andMe Research Team are current or former employees of 23andMe, Inc., and hold stock or stock options in 23andMe. J.B.I. is a patent holder with Biocompatibles UK Ltd. (Framham, Surrey, UK) (Title: Treatment of eye diseases using encapsulated cells encoding and secreting neuroprotective factor and/or anti-angiogenic factor; international patent no. 20120263794) and is in patented a patent application with University of Heidelberg (Heidelberg, Germany) (Title: Agents for use in the therapeutic or prophylactic treatment of myopia or hyperopia; European patent no. 3 070 101). The other authors declare no competing financial interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0127-7.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.C.W.K.
The CREAM Consortium

Tin Aung82,83, Amutha B. Veluchamy82,84, Kathryn P. Burdon58, Harry Campbell36, Li Jia Chen85, Peng Chen83, Wei Chen86, Emily Chew45, Margaret M. Deangelis87, Xiaohu Ding88, Angela Döring66, David M. Evans89,90, Sheng Feng91, Brian Fleck92, Rhys D. Fogarty58, Jeremy R. Fondran43, Maurizio Fossarello93, Xiaobo Guo88,94, Annet E. G. Haarman1,2, Mingguang He23,88, Laura D. Howe90,95, Sarayut Jannmahasatian43, Vishal Jhanji95, Mika Kähönen96, Jaakko Kaprio20,97, John P. Kemp90, Kay-Tee Khaw11, Chiea-Chuen Khor29,83,87,98, Eva Krapohl99, Jean-François Korobelnik100,101, Kristine Lee9, Shi-Ming Li22, Yi Lu56, Robert N. Luben11, Kari-Matti Mäkelä49, George McMahon90, Akira Meguro102, Evelin Mihailov18, Masahiro Miyake16, Nobuhisa Mizuki102, Margaux Morrison87, Vinay Nangia103, Konrad Oexle104, Songhomi Panda-Jonas103, Chi Pui Pang85, Mario Pirastu105, Robert Plomin99, Taina Rantanen77, Maria Schache23, Ilkka Seppälä49, George D. Smith90, Beate St Pourcain90,106, Panayiotis Tatsiou85, J. Willem L. Tideman1,2, Nicholas J. Timpson90, Simona Vaccargiu105, Zoran Vatavuk35, Jie Jin Wang23,24, Ningli Wang22, Nick J. Wareham107, Alan F. Wright33, Shi-Ming Li22, Yi Lu56, Robert N. Luben11, Kari-Matti Mäkelä49, George McMahon90, Akira Meguro102, Evelin Mihailov18, Masahiro Miyake16, Nobuhisa Mizuki102, Margaux Morrison87, Vinay Nangia103, Konrad Oexle104, Songhomi Panda-Jonas103, Chi Pui Pang85, Mario Pirastu105, Robert Plomin99, Taina Rantanen77, Maria Schache23, Ilkka Seppälä49, George D. Smith90, Beate St Pourcain90,106, Panayiotis Tatsiou85, J. Willem L. Tideman1,2, Nicholas J. Timpson90, Simona Vaccargiu105, Zoran Vatavuk35, Jie Jin Wang23,24, Ningli Wang22, Nick J. Wareham107, Alan F. Wright33, Liang Xu22, Maurice K. H. Yap108, Sehyun Yazar74, Shec Ping Yip109, Nagahisa Yoshimura16, Alvin L. Young9, Jing Hua Zhao107 and Xiangtian Zhou86

23andMe Research Team

Michelle Agee7, Babak Alipanahi7, Adam Auton7, Robert K. Bell7, Katarzyna Bryc7, Sarah L. Elson7, Pierre Fontanillas7, David A. Hinds7, Jennifer C. McCreight7, Karen E. Huber7, Aaron Kleinman7, Nadia K. Litterman7, Matthew H. McIntyre7, Joanna L. Mountain7, Elizabeth S. Noblin7, Carrie A. M. Northover7, Steven J. Pitts7, J. Fah Sathirapongsasuti7, Olga V. Sazonova7, Janie F. Shelton7, Suyash Shringarpure7, Chao Tian7, Vladimir Vatic7 and Catherine H. Wilson7

UK Biobank Eye and Vision Consortium

Tariq M. Aslam110, Sarah A. Barman111, Jenny H. Barrett112, Paul N. Bishop110, Peter Blows12, Catey Bunce113, Roxana O. Carare114, Usha Chakravarty115, Michelle Chan12, Sharon Chua12, David Crabb116, Alexander Day12, Parul Desai12, Bal Dhillon117, Andrew D. Dick118, Cathy A. Egan12, Sarah Ennis114, Marcus Fruttiger12, John Gallacher119, David F. Garway-Heath12, Jane Gibson114, Dan M. Gore12, Alison Hardcastle12, Simon P. Harding120, Ruth E. Hogg121, Pearse A. Keane12, Peng Tee Khaw12, Gerassimos Lascaratos12, Andrew Lotery122, Phil J. Luthert12, Tom J. MacGillivray123, Sarah L. Mackie124, Keith R. Martin125, Michelle McGaughey126, Bernadette McGuinness126, Gareth J. McKay126, Martin McKibbin127, Danny Mitry12, Tony Moore12, James E. Morgan26, Zaynah A. Muthy12, Eoin O’Sullivan128, Chris Owen129, Praveen J. Patel12, Euan N. Paterson126, Tunde Peto115, Axel Petzold90, Alicia R. Rudnicka129, Jay E. Self122,131, Sobha Sivaprasad12, David H. W. Steel132, Irene M. Stratton133, Nicholas Strouthidis12, Cathie L. M. Sudlow134, Caroline Thaug12, Dhanes Thomas12, Emanuele Trucco135, Adnan Tufail12, Stephen A. Vernon136, Ananth C. Viswanathan12, Jayne V. Woodside126, Max Yates137, Jennifer L. Y. Yip11 and Yalin Zheng120

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.
Methods

Ethics statement. All human research was approved by the relevant institutional review boards and/or medical ethics committees (listed in Supplementary Note) and conducted according to the Declaration of Helsinki. All CREAM participants provided written informed consent; all 23andMe applicants provided informed consent online and answered surveys according to 23andMe human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, an AAHRPP-accredited institutional review board. The UK Biobank received ethical approval from the National Health Service National Research Ethics Service (reference 11/NW/0382).

Study data. The study populations were participants of the Consortium for Refractive Error and Myopia (CREAM) (41,793 individuals of European ancestry from 26 cohorts (CREAM-EUR) and 11,935 individuals of Asian ancestry from eight studies (CREAM-ASN)) and customers of the 23andMe genetic testing company who provided informed consent for inclusion in research studies (104,293 individuals (two cohorts of individuals with European ancestry, n = 12,128 and n = 92,165, respectively)). All participants included in this analysis from CREAM and 23andMe were 25 years of age or older. Participants with conditions that might alter refraction, such as cataract surgery, laser refractive procedures, retinal detachment surgery, keratoconus, or ocular or systemic syndromes were excluded from the analyses. Recruitment and ascertainment strategies varied by study (Supplementary Table 1a,b and Supplementary Note). Refractive error represented by measurements of refraction and analyzed as spherical equivalent (SphE = refractive error r + 1/2 cylinder refractive error) was the outcome variable for CREAM; myopic refractive error was represented by self-reported AODM for 23andMe.

Genotype calling and imputation. Samples were genotyped on different platforms, and study-specific quality control (QC) measures of the genotyped variants were implemented before association analysis (Supplementary Table 1b). Genotypes were imputed with the appropriate ancestry-matched reference panel for all cohorts from the 1000 Genomes Project (Phase I version 3, March 2012 release) with either minimac3 or IMPUTE2. The metrics for preimputation QC varied among studies, but genotype call-rate thresholds were set at a high level (≥0.95 for both CREAM and 23andMe). These metrics were similar to those of our previous GWAS analyses14,15; details per cohort can be found in Supplementary Table 1b.

GWAS per study. For each CREAM cohort, a single-marker analysis for the phenotype of SphE (in diopters) was carried out with linear regression with adjustment for age, sex and up to the first five principal components. For all non-family-based cohorts, one of each pair of relatives was removed (after detection through either GCTA or identity by sequence (IBS)/identity by descent (IBD) analysis). In family-based cohorts, a score test-based association was used to adjust for within-family relatedness. For the 23andMe participants, Cox proportional hazards analysis testing with AODM as the dependent variable was performed with the following formula: standard error = \(\sqrt{\sum \lambda_{ij} \text{SE}_{ij}^2}\). Analysis was restricted to markers present in the HRC reference panel108. We performed lookups for all independent genetic variants identified in our stage 3 meta-analysis and conditional analysis. For 16 variants not present in UK Biobank, we performed lookups for a surrogate variant in high LD (r^2 > 0.8). When more than one potential surrogate variant was available, the variant in strongest LD focused on variants with a gene-based test implemented in fastBAT41 to the per-variant summary statistics.

Centralized quality control per study. After individual GWAS, all studies were subjected to a centralized round of QC. Quantile-quantile (QQ) plots were generated for each individual cohort in EasyQC24 (Supplementary Fig. 2). All analytical issues discovered during this QC step were resolved per individual cohort.

GWAS meta-analyses. The GWAS meta-analyses were performed in three stages (Supplementary Fig. 1). In stage 1, European (CREAM-EUR, n = 44,192) and Asian (CREAM-ASN, n = 11,935) participants from the CREAM cohort were meta-analyzed separately. Subsequently, all CREAM cohorts (CREAM-ALL) were meta-analyzed. Variants with MAF < 0.1% or imputation quality score < 0.3 (info metric of IMPUTE2 as Song et al.28) were excluded. A fixed-effects inverse-variance-weighted meta-analysis was performed in METAL23,106,107, 1,063 variants clustering in 24 loci (Supplementary Table 2) were genome-wide significant (P = 5.0 × 10^-8). All 37 loci that were previously identified by CREAM and 23andMe by using genotype data inputted to the HapMap II reference panel were replicated (P = 5.0 × 10^-8), and 36 of the 37 were genome-wide significant (P < 5 × 10^-8) (Supplementary Table 2).

In stage 2, a meta-analysis of the two CREAM cohorts (n = 12,318, n = 92,165) was performed with similar filtering but a lower MAF threshold (<0.5%). A total of 5,205 genome-wide significant variants clustered in 112 loci (Supplementary Table 2).

In stage 3, CREAM-ALL and 23andMe samples were combined through a fixed-effects meta-analysis based on P value and direction of effect. In all stages, each genetic variant had to be represented by at least half of the entire study population and represented by at least 13 cohorts in CREAM and one cohort in 23andMe. For SNPs with high heterogeneity (P < 0.05), we also performed a random-effects meta-analysis in METASOFT30. We chose a different weighting scheme because of the differences in effect-size scaling. 23andMe used a less accurate phenotype variable (AODM): the effective sample size for 23andMe was approximately equivalent to the effective sample size of CREAM-ALL (Fig. 3b), and thus we corrected by \(\left(1/\sqrt{\text{MAF}_{23andMe}}\right)\) yielding a final weighting ratio of \(1.1\times\) (ref. 43). Genomewide statistical significance was defined at P < 5.0 × 10^-8 (ref. 44).

All three meta-analysis stages were performed under genomic control. Study-specific and meta-analysis lambda (\(\lambda\)) estimates are shown in Supplementary Fig. 6; to check for confounding biases (for example, cryptic relatedness and population stratification), LD-score intercepts from LD-score regressions per ancestry were constructed (Supplementary Fig. 7). To check the robustness of signals, we ran conventional random-effects models in METASOFT, and fixed-effects models weighted on sample size and on weights estimated from standard error per allele were tested in METAL (Supplementary Table 2 and Supplementary Table 3).

Manhattan (modified version of package ‘qman’) regional, box and forest plots were made in R version 3.2.2 (over 100 variables) and the Venny2.0 (http://bioinfogenomics.pw/) tool was used for the development of the region spanning a 100-kb window of the outermost genome-wide-significant variant of that same region. We annotated all index variants in the web version of ANNOVAR based on UCSC Known Gene Database45. For variants within the coding sequence or 5’ or 3’ UTRs of a gene, that gene was assigned to the index variant (this procedure led to more than one gene being assigned to variants located within the transcription units of multiple overlapping genes). For variants in intergenic regions, the nearest 5’ gene and the nearest 3’ gene were assigned to the variant. Index variants were annotated to functional RNA elements when they were described as such in the UCSC Known Gene Database. We used conservation (phyloP) and prediction tools (SIFT32, MutationTaster33, align GVD34 and PolyPhen-2 (ref. 35)) to predict the pathogenicity of protein-altering exonic variants.

Conditional signal analysis. We performed conditional analysis to identify additional independent signals near the index variant at each locus, using GCTA-COJO32. We transformed the \(z\) scores of the summary statistics to betas with the following formula: standard error = \(1/\sqrt{2}\times\text{SD}_r\times\text{MAF}\) (where \(\text{SD}_r\) is the standard deviation of the reference sample at locus, \(\text{MAF}\) is the minor allele frequency and \(\text{r}\) is the correlation between cases and controls). We performed the GCTA-COJO analysis by using summary-level statistics from the meta-analysis on all cohorts. LD between variants was estimated from S-III–RHI.

Replication in UK Biobank. The UK Ethnogenomics Consortium performed a GWAS of refractive error in 95,505 participants of European ancestry who were 37–73 years of age and had no history of eye disorders13. Refractive error was measured with an autorefractor; SphE was calculated per eye and averaged between the two eyes. To account for relatedness, a mixed-model analysis with BOLT-LMM was used37, including age, sex, genotyping array and the first ten principal components as covariates. Analysis was restricted to markers present in the HRC reference panel46. We performed lookups for all independent genetic variants identified in our stage 3 meta-analysis and conditional analysis. For 16 variants not present in UK Biobank, we performed lookups for a surrogate variant in high LD (r^2 > 0.8). When more than one potential surrogate variant was available, the variant in strongest LD with the index variant was selected. Six variants were available for replication: one variant (rs188159083) was neither present on the array nor was a surrogate available in UK Biobank, and five variants showed evidence of departure from HWE (HWE exact test P < 3.0 × 10^-4).

Post-GWAS analyses. We performed two gene-based tests to identify additional significant genes not found in the single-variant analysis. First, we applied the gene-based test implemented in fastBAT7 to the per-variant summary statistics of the meta-analysis of all European cohorts (23andMe and CREAM-EUR). We used the default parameters (all variants in or within 50 kb of a gene) and focused on variants with a gene-based P < 2 × 10^-4 (Bonferroni correction based on 25,000 genes) and per-variant P > 5 × 10^-4. Second, we applied another gene-based test in EUGENE38, which includes only variants that are eQTLs (Genotype
Tissue Expression (GTEx) data, blood\textsuperscript{(53)}. EUGENE tests a hypothesis predicated on eQTLs as key drivers of the association signal. eQTLs within 50 kb of a gene were included in the test. Genes with EUGENE \(P < 2 \times 10^{-8}\) (and not found in the single variant analysis) were considered significant. Finally, we used functional annotation information from genome-wide-significant loci to reweigh results in lgpaw (version 0.3.64 (ref. 43)). Fgwas incorporates functional annotation (for example, DNsase I–hypersensitive sites in various tissues and 3′-UTR regions) to reweigh data from GWAS and uses a Bayesian model to calculate a posterior probability of association. This approach can identify risk loci that otherwise might not reach the genome-wide-significance threshold in standard GWAS. Details about this approach can be found in the Supplementary Note.

Refractive errors and myopia risk prediction. To assess the risk of the entire range of refractive errors, we computed PGRS values for the population-based RSI-III, using the \(P\) and \(Z\) scores from a meta-analysis on CREAM-ALL and 23andMe, excluding the RSI-III cohorts. Only variants with high imputation quality (IMPUTe info score >0.5 or minimac Rsq >0.8) and MAF >1% were considered. \(P\)-based clumping was performed in PLINK\textsuperscript{(107)}, with an \(r^2\) threshold of 0.2 and a physical-distance threshold of 500 kb, excluding the MHC region. This procedure resulted in a total of 243,938 variants. For each individual in RSI, RSII and RSIII (n = 10,792), PGRS values were calculated with the --score command in PLINK across the following strata of \(P\) thresholds: 5.0 \(\times 10^{-4}\), 5.0 \(\times 10^{-5}\), 5.0 \(\times 10^{-6}\), 5.0 \(\times 10^{-7}\), 0.005, 0.01, 0.05, 0.1, 0.5, 0.8 and 1.0. The proportion of variance explained by each PGRS model was calculated as the difference in the \(R^2\) between two regression models: one in which SphE was regressed on age, sex and the first five principal components, and the other also including the PGRS as an additional covariate. Subsequently, areas under the receiver operating characteristic curve were mapped for SphE (\(\geq 3\) s.d.) vs. hyperopia (\(\geq 3\) s.d.).

Genetic correlation between ancestries. We used Popcorn\textsuperscript{(111)} to investigate ancestry-related differences in the genetic architecture of refractive error and myopia. Popcorn takes summary GWAS statistics from two populations and LD information from ancestry-matched reference panels, and computes genetic correlations by implementing a weighted likelihood function that accounts for the inflation of \(Z\) scores due to LD. Pairwise analyses were carried out by using the GWAS summary statistics from 23andMe (n = 104,292), CREAM-EUR (n = 44,192) and CREAM-EAS (n = 9,826) meta-analyses. Only SNPs with MAF \geq 5% were included, thus resulting in a final set of 3,625,602 SNPs for analyses involving 23andMe and 3,642,928 SNPs for the CREAM-EUR vs. CREAM-EAS analysis. Reference panels were constructed with genotype data from 503 European and 504 East Asian individuals sequenced as part of the 1000 Genomes Project (release 2 May 2013; see URLs). The reference-panel VCF files were filtered in PLINK\textsuperscript{(110)} to remove indels, strand-ambiguous variants, variants without an ‘rs’ ID prefix and variants located in the MHC region on chromosome 6 (chromosome 6: 250000000–335000000; build 37).

Analysis between phenotypes. To evaluate the consistency of genotypic effects across studies that used different phenotype definitions, we compared effect sizes from GWAS studies of either SphE or AODM in Europeans, i.e., CREAM-EUR (\(n = 44,192\)) or 23andMe (\(n = 104,293\)), respectively. Marker-wise additive genetic effect sizes (in diploters per copy of the risk allele) for SphE were compared against those (in units log(HR) per copy of the risk allele) for AODM. Data were visualized with R. Genetic correlation between the two phenotypes SphE and AODM was calculated through LD-score regression. This analysis included all common SNPs (MAF >0.01) present in HapMap3.

Evidence of functional involvement. To rank genes according to biological plausibility, we scored annotated genes according to our own findings and published reports of a potential functional role in refractive error. Points were assigned for each gene on the basis of ten categories (details on the methodology per category are provided in Supplementary Note): internal replication of index genetic variants in the individual cohort GWAS analyses through Bonferroni correction (CREAM-ASN, CREAM-EUR and 23andMe; \(P_{\text{Bonferroni}} < 1.19 \times 10^{-5}\)); evidence of eQTL from FUMA\textsuperscript{(112)} analysis and extensive lookups in GTEx; evidence of expression in the eye in developmental ocular tissues; evidence of expression in the eye in adult ocular tissues; presence of an eye phenotype in knockout mice (Mouse Genome Informatics and International Mouse Phenotyping Consortium databases); presence of an eye phenotype in humans (OMIM; see URLs, DisGeNET\textsuperscript{(113)}); location in a functional region of a gene (wANNOVAR; see URLs); presence of the gene in a significant enriched functional pathway with FDR <0.05 (DEPICT\textsuperscript{(114)}); presence of the gene in the gene pathway analysis of DEPICT with FDR <0.05; and presence of the gene in the canonical pathway analysis of IPA (see URLs). Furthermore, we performed a systematic search for each gene to assess its potential as a drug target (SuperTarget\textsuperscript{(115)}, STITCH\textsuperscript{(116)}, DrugBank\textsuperscript{(117)} and Pharmafacts\textsuperscript{(118)}). All information derived from this study and the literature was used to annotate genes to retinal cell types.

Genetic pleiotropy. To investigate the overlap of genes with other common traits, we performed a lookup in the GWAS catalog by using FUMA. Multiple-testing correction (i.e., Benjamini-Hochberg) was performed. Traits were significantly associated when adjusted \(P \leq 0.05\), and the number of genes that overlapped with the GWAS-catalog gene sets was 22.

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

Data availability. The summary statistics of the stage 3 meta-analysis are included in Supplementary Data 3. To protect the privacy of the participants in our cohorts, further summary statistics of stage 1 (CREAM) and stage 2 (23andMe) will be available upon reasonable request. Please contact c.c.w.klaver@erasmusmc.nl or apply.research@23andMe.com (23andMe) for more information and to access the data.

References

96. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genome-wide association testing through pre-phasing. Nat. Genet. 44, 955–959 (2012).
97. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
98. Chen, W. M. & Abecasis, G. R. Family-based association tests for genome-wide association scans. Bioinformatics 26, 2190–2191 (2010).
99. Winker, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. 9, 1192–1212 (2014).
100. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. Bioinformatics 26, 2336–2337 (2010).
101. Zaykin, D. V. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J. Evol. Biol. 24, 1836–1841 (2011).
102. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. Genet. Epidemiol. 32, 227–234 (2008).
103. Puim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
104. Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. 10, 1556–1566 (2015).
105. Cooper, G. M. et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 15, 901–913 (2005).
106. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods 7, 575–576 (2010).
107. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. 47, 284–290 (2015).
108. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. 48, 1279–1283 (2016).
109. Consortium, G. T., GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multisite tissue gene regulation in humans. Science 348, 648–660 (2015).
110. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
111. Bauer-Mehren, A., Rautschka, M., Sanz, F. & Furlong, L. I. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics 26, 2924–2926 (2010).
112. Günther, S. et al. SuperTarget and Matador: resources for exploring drug-target relationships. Nucleic Acids Res. 36, D919–D922 (2008).
113. Kuhn, M. et al. STITCH 4: integration of protein-chemical interactions with user data. Nucleic Acids Res. 42, D401–D407 (2014).
114. Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 34, D688–D692 (2006).
115. Whirl-Carrillo, M. et al. Pharmacogenomics knowledge for personalized medicine. Clin. Pharmacol. Ther. 92, 414–417 (2012).
Experimental design

1. Sample size
   Describe how sample size was determined.
   Our strategy aimed to create the largest possible sample size for the meta-analysis and we initially included practically all existing population studies with genetic and refractive error data in our analysis. Furthermore, for the replication analysis, we used the summary statistics of the GWAS from the UKEV consortium based on refractive error. We performed a power calculation using G*Power 3.1.9.2 in order to check the power of the sample size of this cohort (n= 95,505): the two-sided linear multiple regression t-test with a mean effect of 0.03, an alpha of 0.000299 (0.05/167) and at least 80% power, the appropriate sample size for replication should comprise at least 669 participants. The UKEV cohort is the largest and only other independent cohort known in the world with this similar accurate phenotype.

2. Data exclusions
   Describe any data exclusions.
   Every cohort removed participants with conditions that could alter refraction, such as cataract surgery, laser refractive procedures, retinal detachment surgery, keratoconus as well as ocular or systemic syndromes.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   There are no other existing large studies to replicate our findings to date. We performed internal and independent replications. We found significant overlap in the internal replications: all 25 loci identified at Stage 1 (CREAM) replicated in Stage 2 (23andMe; pBonferroni 2.00 x 10^-3). Vice versa, 29 (25.9%) of the loci identified at Stage 2 replicated in Stage 1 (pBonferroni 4.46 x 10^-4), an expected proportion given the lower statistical power in CREAM. Furthermore, we replicated in an independent cohort consisting of 95,505 participants. In the GWAS on refractive error performed by the UK Biobank Eye & Vision Consortium, we replicated 86% of all independent loci.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Randomization was not relevant to our GWAS meta-analysis study; we performed an overall meta-analyses of all available data.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Blinding was not relevant to our study; our analysts only had access to summary statistics of GWAS analyses.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|
| ☐  | ☒         | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐  | ☒         | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly. |
| ☐  | ☒         | A statement indicating how many times each experiment was replicated |
| ☐  | ☒         | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐  | ☒         | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐  | ☒         | The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted |
| ☐  | ☒         | A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐  | ☒         | Clearly defined error bars |

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

- R version 3.2.3 (packages: qqman, ggplot2, metafor); Minimac, IMPUTE (imputations); EasyQC version: 9.0 (quality control); METAL 2011-03-25 release (GWAS meta-analyses); LocusZoom (regional plots); LDSC https://github.com/bulik/ldsc (LD score regression); GCTA64 version 1.26.0 (conditional analyses); fastbat, EUGENE, fgwas (post GWAS analyses); PLINK v1.9 (clumping for PGRS); Popcorn https://github.com/brielin/Popcorn (ancestry-related differences); FUMA (eQTLS & GWAS catalogue look up); DEPICT v1 release 194, Cytoscape version 3.4.0, IPA (pathway analysis); Polyphen [link]; SIFT [link]; Mutation Taster [link]; METASOFT v2.0.1 (Random Effects meta-analyses)

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

- No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

- No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

- No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

- No eukaryotic cell lines were used.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

- No commonly misidentified cell lines were used.
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.
   No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.
   All participants included in this analysis from CREAM and 23andMe were aged 25 years or older. Participants with conditions that could alter refraction, such as cataract surgery, laser refractive procedures, retinal detachment surgery, keratoconus as well as ocular or systemic syndromes were excluded from the analyses. All relevant information on the study participants, including mean age, gender, and refractive error is stated in Supplementary Table 1a,b. No individual genotype data are shared. Refractive error represented by measurements of refraction and analyzed as spherical equivalent (SphE = spherical refractive error + 1/2 cylinder refractive error) was the outcome variable for CREAM; myopic refractive error represented by self-reported age of diagnosis of myopia (AODM) for 23andMe. For each CREAM cohort, a single marker analysis for the SphE (in diopters) phenotype was carried out using linear regression adjusting for age, sex and up to the first five principal components.