

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

Milly S. Tedja^{1,2,80}, Robert Wojciechowski^{3,4,5,80}, Pirro G. Hysi^{6,80}, Nicholas Eriksson^{7,80}, Nicholas A. Furlotte^{7,80}, Virginie J. M. Verhoeven^{1,2,8,80}, Adriana I. Iglesias^{1,2,8}, Magda A. Meester-Smoor^{1,2}, Stuart W. Tompson⁹, Qiao Fan¹⁰, Anthony P. Khawaja^{11,12}, Ching-Yu Cheng^{10,13}, René Höhn^{14,15}, Kenji Yamashiro¹⁶, Adam Wenocur¹⁷, Clare Grazal¹⁷, Toomas Haller¹⁸, Andres Metspalu¹⁸, Juho Wedenoja^{19,20}, Jost B. Jonas^{21,22}, Ya Xing Wang²², Jing Xie²³, Paul Mitchell²⁴, Paul J. Foster¹², Barbara E. K. Klein⁹, Ronald Klein⁹, Andrew D. Paterson²⁵, S. Mohsen Hosseini²⁵, Rupal L. Shah²⁶, Cathy Williams²⁷, Yik Ying Teo^{28,29}, Yih Chung Tham¹³, Preeti Gupta³⁰, Wanting Zhao^{10,31}, Yuan Shi³¹, Woei-Yuh Saw³², E-Shyong Tai²⁹, Xue Ling Sim²⁹, Jennifer E. Huffman³³, Ozren Polasek³⁴, Caroline Hayward³³, Goran Bencic³⁵, Igor Rudan³⁶, James F. Wilson^{33,36}, CREAM³⁷, 23andMe Research Team³⁷, UK Biobank Eye and Vision Consortium³⁷, Peter K. Joshi³⁶, Akitaka Tsujikawa¹⁶, Fumihiko Matsuda³⁸, Kristina N. Whisenhunt⁹, Tanja Zeller³⁹, Peter J. van der Spek⁴⁰, Roxanna Haak⁴⁰, Hanne Meijers-Heijboer^{41,42}, Elisabeth M. van Leeuwen^{1,2}, Sudha K. Iyengar^{43,44,45}, Jonathan H. Lass^{43,44}, Albert Hofman^{2,46,47}, Fernando Rivadeneira^{2,47,48}, André G. Uitterlinden^{2,47,48}, Johannes R. Vingerling¹, Terho Lehtimäki^{49,50}, Olli T. Raitakari^{51,52}, Ginevra Biino⁵³, Maria Pina Concas⁵⁴, Tae-Hwi Schwantes-An^{4,55}, Robert P. Igo Jr⁴³, Gabriel Cuellar-Partida⁵⁶, Nicholas G. Martin⁵⁷, Jamie E. Craig⁵⁸, Puya Gharakhani⁵⁶, Katie M. Williams⁶, Abhishek Nag⁵⁹, Jugnoo S. Rahi^{12,60,61}, Phillippa M. Cumberland⁶⁰, Cécile Delcourt⁶², Céline Bellenguez^{63,64,65}, Janina S. Ried⁶⁶, Arthur A. Bergen^{41,67,68}, Thomas Meitinger^{69,70}, Christian Gieger⁶⁶, Tien Yin Wong^{71,72}, Alex W. Hewitt^{23,73,74}, David A. Mackey^{23,73,74}, Claire L. Simpson^{4,75}, Norbert Pfeiffer¹⁵, Olavi Pärssinen^{76,77}, Paul N. Baird²³, Veronique Vitart³³, Najaf Amin², Cornelia M. van Duijn², Joan E. Bailey-Wilson⁴, Terri L. Young⁹, Seang-Mei Saw^{29,78}, Dwight Stambolian¹⁷, Stuart MacGregor⁵⁶, Jeremy A. Guggenheim^{26,81}, Joyce Y. Tung^{7,81}, Christopher J. Hammond^{6,81} and Caroline C. W. Klaver^{1,2,79,81*}

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^{1–4}. The prevalence of myopia is also rising throughout Western Europe and the United States, affecting ~50% of young adults in these regions^{5,6}. Although refractive errors can be optically corrected, even at moderate values they carry substantial risk of ocular complications with high economic burden^{7–9}. 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^{10,11}. At the other extreme, high hyperopia predisposes individuals to strabismus, amblyopia and angle-closure glaucoma^{10,12}.

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 work³. Recent research has identified many genetic variants for refractive errors, myopia and axial length^{13–25}. Two large studies—the International Consortium for Refractive Error and Myopia (CREAM)²⁶ and the personal genomics company 23andMe, Inc.^{17,27}—have provided the most comprehensive results²⁸.

Given that only 3.6% of the variance of the refractive-error trait was explained by the identified genetic variants²⁶, 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^{26,27} (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 I reference panel (version 3, March 2012 release²⁹) 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,b, 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 regression³⁰ (stage 1 and 2 inflation factors (GC) <1.1 and LD-score regression intercepts ($LDSC_{intercept}$) 0.892–1.023; Supplementary Table 6 and Supplementary Figs. 6 and 7). At stage 3, we observed genomic inflation (λ_{GC} =1.129; Supplementary Fig. 6), probably because of true polygenicity rather than population stratification or cryptic relatedness³¹. $LDSC_{intercept}$ 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-COJO³² 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 ($n_{RSI-III}$ =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)³³ (Methods). Six out of the 167 variants were not considered for replication analysis. One of these five variants (rs3138141, *RDH5*) was identified previously and therefore still considered a refractive-error risk variant^{26,27}. 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 ($P < 3.0 \times 10^{-4}$; 21.1%); another 12 showed nominal evidence for replication ($0.05 < P < 3.0 \times 10^{-4}$; 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 $\log(\text{hazard ratio}(\text{HR}))$ 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 $\log(\text{HR})$ 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; $P = 2.1 \times 10^{-159}$), 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)³⁴. 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 $MAF \leq 0.05$: 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 SIFT³⁶. *PDE11A* is presumed to play a role in tumorigenesis, brain function and inflammation³⁷. 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, SIFT³⁸ and align GVD^{39,40}. 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, $P = 4.21 \times 10^{-84}$) 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 ($P = 2.28 \times 10^{-65}$).

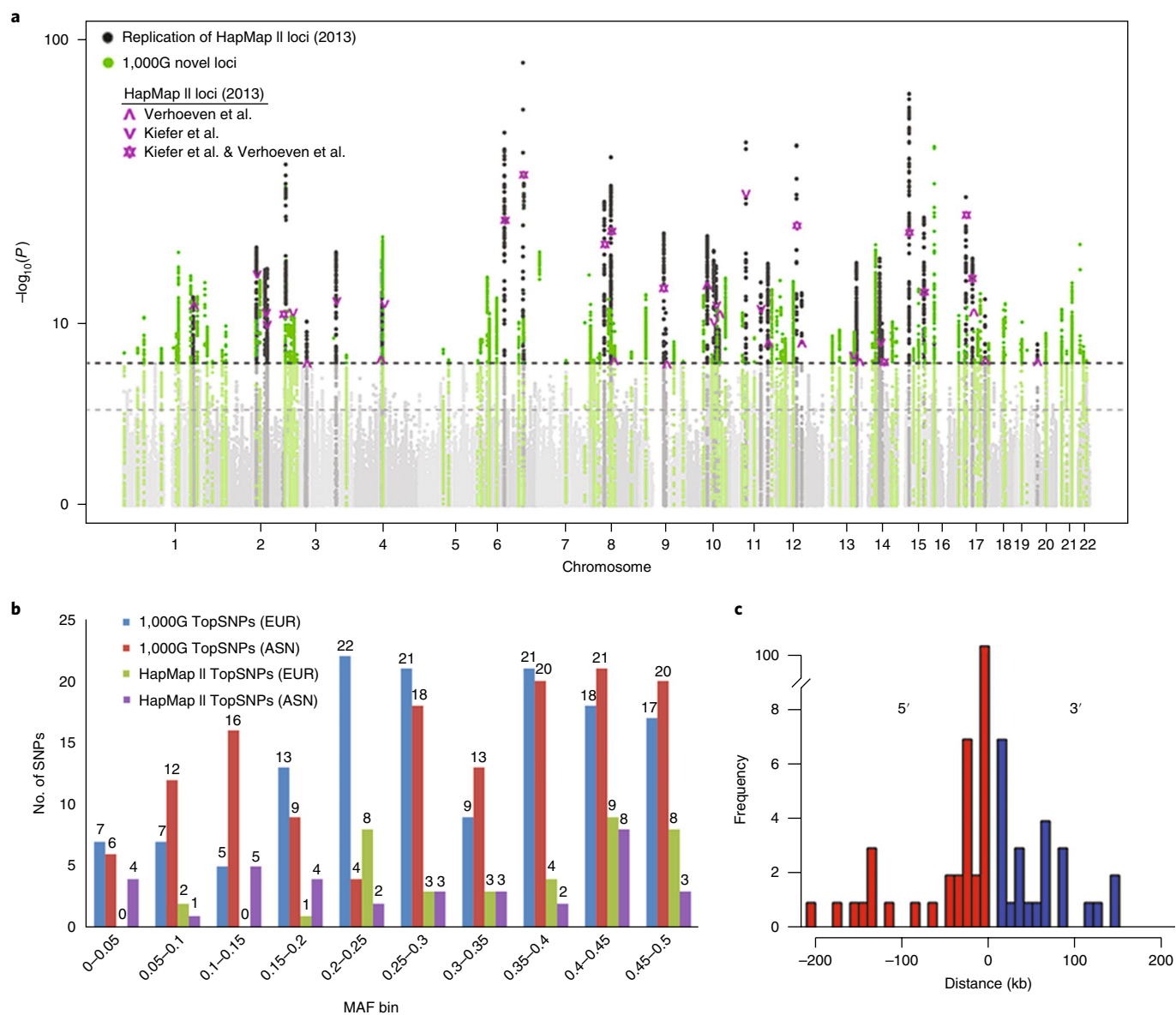


Fig. 1 | GWAS meta-analysis identifies 140 loci for refractive error (stage 3). **a**, Meta-analysis of genome-wide single-variant analyses for >10 million variants in 160,420 CREAM and 23andMe participants (stage 3). Shown is a Manhattan plot depicting P for association, highlighting newly identified ($P < 5 \times 10^{-8}$; green) and known (dark gray) refractive-error loci previously found by using HapMap II imputations from Kiefer et al.²⁷ and Verhoeven et al.²⁶ (Table 1). The horizontal lines indicate suggestive significance ($P = 1 \times 10^{-5}$) or genome-wide significance ($P = 5 \times 10^{-8}$). 1000G, 1000 Genomes Project. **b**, MAFs of the 140 discovered index variants based on 1000G (blue, Europeans; red, Asians) to the MAFs of the previously found genetic variants based on HapMap II (green, Europeans; purple, Asians). An increase was observed in genetic variants found across all MAF bins, including the lower MAF bins. **c**, Annotation of the 167 loci to genes in wANNOVAR. Shown are the distances between index variants from the nearest gene and its gene on the 5' and/or 3' site. Most index variants (84%) were at a distance of less than 50 kb up- or downstream from the annotated gene.

Post-GWAS analyses. We performed two gene-based tests, fastBAT⁴¹ and EUGENE⁴², and applied a functional enrichment approach with fgwas⁴³ (Methods). With fastBAT, we identified 13 genes at $P < 2.0 \times 10^{-6}$, one of which (*CHD7*) had been identified previously^{26,27}. 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)⁴⁴ 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

Replication of the HapMap II index variants for refractive error per locus in the stage 3 meta-analysis

SNP	Chromosome	Position	Nearest loci and gene(s)	Effect allele	Other allele	EAF EUR	EAF ASN	Z score	Direction	P value	Het /sq	Het P value	Sample size (n)	P-value replication
rs12193446	6	129820038	BC035400, LAMA2	A	G	0.906	NA	-19.43	--	4.21 × 10 ⁻⁸⁴	16.8	5.72 × 10 ⁻¹⁵	150,269	4.60 × 10 ⁻¹⁰⁶
rs524952	15	35005586	GOLGA8B, GJD2	A	T	0.475	0.507	-17.08	--	2.28 × 10 ⁻⁶⁵	67.2	0.015	160,150	1.60 × 10 ⁻¹⁰³
rs7744813	6	73643289	KCNQ5	A	C	0.591	0.602	-14.56	--	5.43 × 10 ⁻⁴⁸	35	0.001	160,091	1.00 × 10 ⁻⁷⁵
rs11602008	11	40149305	LRR4C	A	T	0.822	0.749	13.98	++	2.12 × 10 ⁻⁴⁴	22.5	1.56 × 10 ⁻¹⁰	157,505	2.90 × 10 ⁻⁴⁷
rs3138141	12	56115778	BLOC1S1-RDH5, RDH5	A	C	0.214	0.147	13.8	++	2.46 × 10 ⁻⁴³	3.2	5.05 × 10 ⁻⁷	157,531	2.30 × 10 ⁻⁵⁶
rs10500355	16	7459347	RBFOX1	A	T	0.354	0.133	-13.73	--	6.49 × 10 ⁻⁴³	9.1	2.93 × 10 ⁻⁷	160,139	2.50 × 10 ⁻⁴⁸
rs72621438	8	60178580	SNORA51, CA8	C	G	0.642	0.609	-13.14	--	2.03 × 10 ⁻³⁹	38.4	0.006	160,128	1.80 × 10 ⁻⁴⁹
rs1550094	2	233385396	CHRNA3, PRSS56	A	G	0.701	0.705	12.74	++	3.64 × 10 ⁻³⁷	26.3	0.003	159,422	4.10 × 10 ⁻⁵⁹
rs2908972	17	11407259	SHISA6	A	T	0.415	0.484	-11.13	--	9.46 × 10 ⁻²⁹	23	0.254	160,123	6.10 × 10 ⁻²⁹
rs7829127	8	40726394	ZMAT4	A	G	0.792	0.897	-10.91	--	1.02 × 10 ⁻²⁷	15.9	2.77 × 10 ⁻⁴	160,132	3.10 × 10 ⁻²²
rs6495367	15	79375347	RASGRF1	A	G	0.408	0.399	-10.2	--	1.95 × 10 ⁻²⁴	0	0.667	160,144	7.20 × 10 ⁻³⁷
rs11145465	9	71766593	TJP2	A	C	0.212	NA	-9.55	--	1.35 × 10 ⁻²¹	46.3	0.1722	153,174	1.00 × 10 ⁻¹⁰
rs1649068	10	60304864	BICC1	A	C	0.475	0.504	-9.44	--	3.77 × 10 ⁻²¹	0	0.712	160,144	7.50 × 10 ⁻¹¹
rs7692381	4	81903049	CFAP299, BMP3	A	G	0.763	0.63	9.4	++	5.55 × 10 ⁻²¹	0	0.013	160,134	7.50 × 10 ⁻¹³
rs56075542	2	146882415	BC040861, PABPC1P2	T	G	0.552	0.472	-8.99	--	2.39 × 10 ⁻¹⁹	13.9	0.001	159,478	1.30 × 10 ⁻¹⁸
rs7895108	10	79061458	KCNMA1	T	G	0.351	0.118	-8.87	--	7.56 × 10 ⁻¹⁹	32.8	0.021	160,140	1.10 × 10 ⁻²⁷
rs7624084	3	141093285	ZBTB38	T	C	0.568	0.633	-8.81	--	1.24 × 10 ⁻¹⁸	18.5	0.018	160,151	6.50 × 10 ⁻¹⁷
rs62070229	17	31227593	MYO1D, TME0M98	A	G	0.807	0.874	8.58	++	9.64 × 10 ⁻¹⁸	0	0.416	156,570	1.30 × 10 ⁻¹⁸
rs2855530	14	54421917	BMP4	C	G	0.507	0.474	-8.58	--	9.87 × 10 ⁻¹⁸	41.7	0.19	160,092	4.80 × 10 ⁻²²
rs7662551	4	80537638	LOC100506035, PCAT4	A	G	0.723	0.558	8.53	++	1.47 × 10 ⁻¹⁷	19.4	0.265	160,147	6.00 × 10 ⁻¹²
rs9517964	13	100717833	ZIC2, PCCA	T	C	0.589	0.786	8.42	++	3.68 × 10 ⁻¹⁷	0	0.02	160,121	3.40 × 10 ⁻²⁰
rs1954761	11	105596885	GRIA4	T	C	0.371	0.377	-8.4	--	4.57 × 10 ⁻¹⁷	0	0.911	160,122	1.20 × 10 ⁻¹⁶
rs745480	10	85986554	LRR12, LRIT1	C	G	0.511	0.418	8.31	++	9.26 × 10 ⁻¹⁷	67.3	0.081	159,504	8.20 × 10 ⁻¹⁸
rs2573081	2	178828507	PDE11A	C	G	0.524	0.538	8.21	++	2.18 × 10 ⁻¹⁶	47.6	0.167	160,126	1.60 × 10 ⁻²⁹
rs17428076	2	172851936	HATI, METAP1D	C	G	0.768	0.854	-8.18	--	2.77 × 10 ⁻¹⁶	0	0.003	160,151	7.50 × 10 ⁻⁸
rs2155413	11	84634790	DLG2	A	C	0.482	0.655	-7.76	--	8.85 × 10 ⁻¹⁵	0	2.99 × 10 ⁻⁴	159,504	1.10 × 10 ⁻¹⁷
rs11178469	12	71275137	PTPRR	T	C	0.752	0.638	-7.4	--	1.33 × 10 ⁻¹³	0	0.6989	160,139	2.60 × 10 ⁻⁰⁴
rs1858001	1	207488004	C4BPA, CD55	C	G	0.676	0.415	7.28	++	3.45 × 10 ⁻¹³	59.6	0.02	160,149	6.70 × 10 ⁻²⁰
rs4793501	17	68718734	KCNJ2, BC039327	T	C	0.575	0.444	-7.21	--	5.53 × 10 ⁻¹³	0	0.592	160,150	3.70 × 10 ⁻¹²
rs7042950	9	77149837	RORB	A	G	0.732	0.392	6.8	++	1.07 × 10 ⁻¹¹	0	0.912	160,153	2.90 × 10 ⁻¹⁸

Continued

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)

Replication of the HapMap II index variants for refractive error per locus in the stage 3 meta-analysis

SNP	Chromosome	Position	Nearest loci and gene(s)	Effect allele	Other allele	EAF EUR	EAF ASN	Z score	Direction	P value	Het Isq	Het P value	Sample size (n)	P-value replication
rs4687586	3	53837971	CACNA1D	C	G	0.691	NA	-6.55	--	5.86×10^{-11}	0	0.605	150,217	1.60×10^{-8}
rs2753462	14	60850703	JB175233, C14orf39	C	G	0.296	0.568	-6.49	--	8.37×10^{-11}	73.9	0.05	157,352	2.00×10^{-15}
rs837323	13	101175664	PCCA	T	C	0.512	0.762	6.32	++	2.65×10^{-10}	35.6	0.213	160,142	5.30×10^{-16}
rs17382981	10	94953258	CYP26A1, MYOF	T	C	0.417	0.19	-6.31	--	2.72×10^{-10}	67.9	0.077	155,332	4.10×10^{-7}
rs79266634	16	7309047	RBF0X1	C	G	0.093	0.115	-5.93	--	3.00×10^{-9}	0	0.561	156,268	1.50×10^{-8}
rs235770	20	6761765	BMP2	T	C	0.372	0.388	-5.93	--	3.11×10^{-9}	0	0.547	157,521	4.80×10^{-11}
rs36024104	14	42294993	LRFN5	A	G	0.823	NA	9.09	++	9.86×10^{-20}	15.9	0.01414	152,585	2.20×10^{-12}
rs1556867	1	164213686	5S_rRNA, PBX1	T	C	0.264	0.494	-8.81	--	1.29×10^{-18}	71.1	0.06266	160,155	4.20×10^{-17}
rs2225986	1	200311910	LINC00862	A	T	0.381	0.169	-7.96	--	1.68×10^{-15}	40.2	0.196	160,152	7.50×10^{-17}
rs1207782	6	22059967	LINC00340	T	C	0.577	0.265	-7.92	--	2.47×10^{-15}	0	0.8946	160,149	4.90×10^{-13}
rs72826094	10	114801488	TCF7L2	A	T	0.799	0.838	7.88	++	3.20×10^{-15}	64.5	0.09323	156,825	4.90×10^{-2}
rs297593	2	157363743	GPD2	T	C	0.286	0.257	-7.82	--	5.45×10^{-15}	0	0.5285	159,461	7.80×10^{-11}
rs5442	12	6954864	GNB3	A	G	0.068	NA	-7.82	--	5.48×10^{-15}	8.8	0.03693	146,217	1.20×10^{-33}
rs10880855	12	46144855	ARID2	T	C	0.507	0.464	-7.78	--	7.35×10^{-15}	0	0.9683	160,144	4.80×10^{-8}
rs2150458	21	47377296	PCBP3, COL6A1	A	G	0.455	0.641	7.74	++	1.04×10^{-14}	55.7	0.1329	160,151	1.80×10^{-13}
rs12898755	15	63574641	APHIB	A	G	0.245	0.456	7.53	++	4.98×10^{-14}	7.9	0.2974	159,506	1.40×10^{-16}
rs7122817	11	117657679	DSCAML1	A	G	0.507	0.662	7.51	++	5.73×10^{-14}	73.8	0.05077	160,147	1.10×10^{-10}
rs10511652	9	18362865	SH3GL2, ADAMTSL1	A	G	0.416	0.445	7.36	++	1.91×10^{-13}	44.8	0.1782	160,149	3.50×10^{-18}
rs11101263	10	49414181	FRMPD2	T	C	0.258	0.105	-7.33	--	2.33×10^{-13}	0	0.3477	160,155	2.20×10^{-13}
rs11118367	1	219790221	LYPLAL1	T	C	0.482	0.630	-7.29	--	3.16×10^{-13}	0	0.8576	160,141	1.20×10^{-13}
rs9395623	6	50757699	TFAP2D, TFAP2B	A	T	0.315	0.381	7.25	++	4.16×10^{-13}	0	0.9579	160,151	2.20×10^{-10}
rs284816	8	53362145	ST18, FAM150A	A	G	0.163	0.198	-7.21	--	5.52×10^{-13}	0	0.9242	160,140	1.60×10^{-8}
rs12965607	18	47391025	MYO5B	T	G	0.857	0.923	7.07	++	1.52×10^{-12}	20.8	0.01674	157,604	8.10×10^{-16}
rs7747	4	80827062	ANTXR2	T	C	0.202	0.093	7.03	++	2.05×10^{-12}	5.4	0.01267	150,327	7.70×10^{-16}
rs12451582	17	54734643	NOG, C17orf67	A	G	0.369	0.308	7.02	++	2.22×10^{-12}	0	0.5925	160,155	8.80×10^{-18}
rs80253120	17	14138507	CDRT15	T	C	0.626	0.723	6.97	++	3.25×10^{-12}	58.6	0.12	156,054	7.20×10^{-11}
rs7968679	12	9313304	PZP	A	G	0.700	0.894	6.95	++	3.65×10^{-12}	0	0.01951	160,076	4.20×10^{-10}
rs11202736	10	90142203	RNL5	A	T	0.717	0.762	-6.92	--	4.53×10^{-12}	0	0.4007	160,150	9.40×10^{-7}
rs72655575	8	60556509	SNORA51, CA8	A	C	0.201	0.124	6.87	++	6.54×10^{-12}	0	0.8811	156,566	7.10×10^{-7}
rs1790165	11	131928971	NTM	A	C	0.411	0.283	6.85	++	7.17×10^{-12}	0	0.003708	160,131	1.80×10^{-10}
rs511217	11	30029948	METTL15, KCNA4	A	T	0.738	0.729	-6.79	--	1.10×10^{-11}	0	0.3626	160,143	1.40×10^{-17}

We identified 140 loci for refractive error with genome-wide significance ($P < 5 \times 10^{-8}$) on the basis of the meta-analyses of the genome-wide single-variant linear regressions performed in 160,420 participants of mixed ancestries (CREAM-ASN, CREAM-EUR and 23andMe). Shown are the replication of the previously found loci from HapMap II and a subset of the new loci with the smallest P values. For each locus, represented by an index variant (the variant with the smallest P value in that locus), effect allele, other allele, effect-allele frequencies per ancestry (EAF ASN and EAF EUR), effect size (Z score), direction of the effect (direction), the P value, heterogeneity / square (Het Isq), heterogeneity P value (Het P value), sample size (n) and P value of the replication in UK Biobank are shown (full table in Supplementary Table 2). ASN, Asian; EUR, European; GWS, genome wide significant; NA, not applicable.

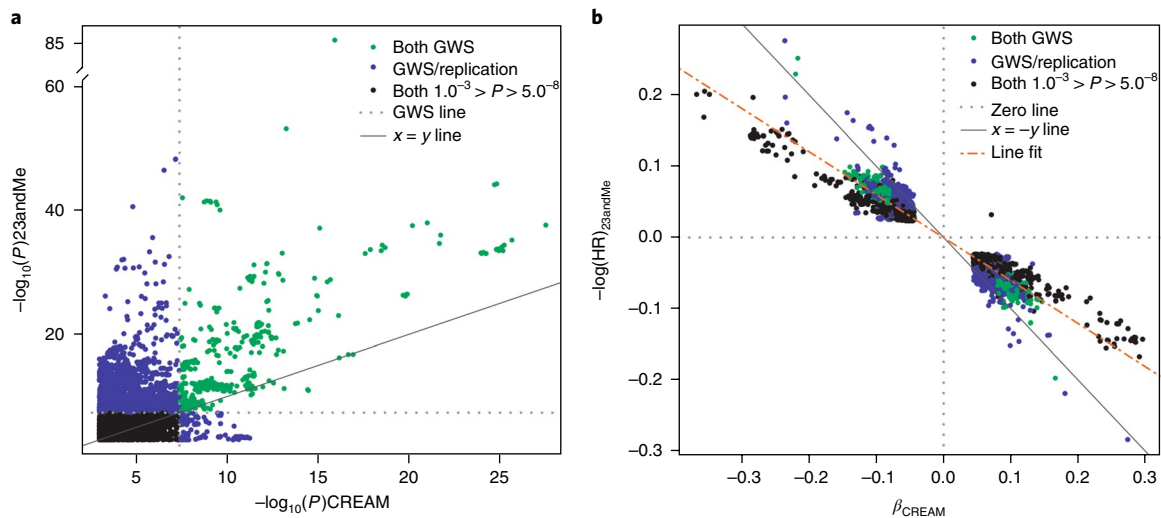


Fig. 2 | Correlation of statistical significance and effect size of SNPs on the basis of SpHE in diopters and AODM in years. a, P comparison of all genetic variants with $P < 1.0 \times 10^{-3}$ ($n = 7,249$) between CREAM meta-analysis (stage 1) and 23andMe (stage 2) meta-analysis. Shown is the overlap (green) and the difference (purple) in P signals per cohort for genetic variants. Purple genetic variants are only genome wide significant (GWS) or 23andMe. Black, genetic variants with P between 5.0×10^{-8} and 1.0×10^{-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.

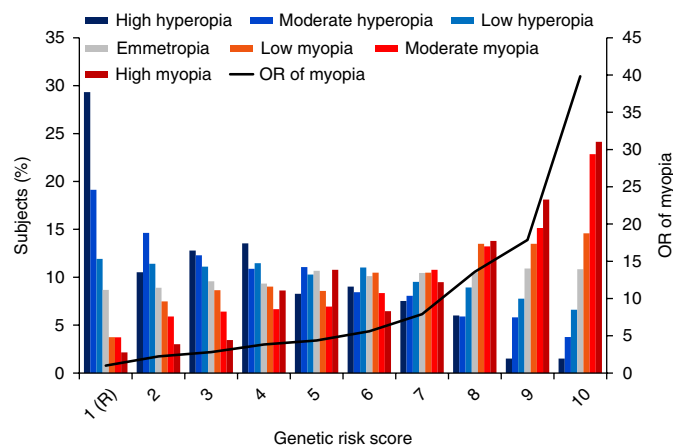


Fig. 3 | Risk of refractive error per decile of polygenic risk score (Rotterdam Study I-III, $n = 10,792$). Distribution of refractive error in subjects from RSI-III ($n = 10,792$) as a function of the optimal polygenic risk score (including 7,303 variants at $P \leq 0.005$ explaining 7.8% of the variance 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 ≤ -3 diopters; low myopia, SpHE > -3 diopters and < -1.5 diopters; emmetropia, SpHE ≥ -1.5 diopters and ≤ 1.5 diopters; low hyperopia, SpHE > 1.5 diopters and < 3 diopters; moderate hyperopia, SpHE ≥ 3 diopters and < 6 diopters; high hyperopia, SpHE ≥ 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 estimations⁴⁵.

Transancestral 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 regression⁴⁶. 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 Popcorn⁴⁷ (Methods). Two genetic correlation metrics were calculated: (i) a genetic-effect correlation (ρ_{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 (ρ_{gi}) that estimated the correlation in variance-normalized SNP effect sizes between the two ancestry groups (Table 2). Estimates of ρ_{ge} were high between Europeans and Asians, but were significantly different from 1 (0.79 and 0.80, respectively, at $P < 1.9 \times 10^{-6}$; Table 2), thus indicating a clear genetic overlap but a difference in per-allele effect size. Estimates of ρ_{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 DEPICT⁴⁸ to perform a gene set enrichment analysis, a tissue-type enrichment analysis and a gene prioritization analysis, on all variants with $P < 1.00 \times 10^{-5}$ 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; $P = 1.79 \times 10^{-7}$). 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

Table 2 | Genetic correlation for refractive error between Europeans and East Asians

Sample 1	Sample 2	Genetic effect correlation (ρ_{ge}) ^a	Standard error ρ_{ge}	P value ρ_{ge}	Genetic impact correlation (ρ_{gi}) ^a	Standard error ρ_{gi}	P value ρ_{gi}
EUR CREAM	EAS CREAM	0.804	0.041	1.83×10^{-6}	0.888	0.061	0.065
EUR 23andMe	EAS CREAM	0.788	0.041	2.48×10^{-7}	0.865	0.054	0.014

Abbreviations: EUR, European; EAS, East Asian. ^aP value relates to a test of the null hypothesis that $\rho_{ge}=1$ or $\rho_{gi}=1$. We calculated the genetic correlation of effect (ρ_{ge}) and impact (ρ_{gi}) by using Popcorn to compare the genetic associations between Europeans (CREAM-EUR, $n=44,192$; 23andMe, $n=104,292$) and East Asians (CREAM-ASN, $n=9,826$). Reference panels for Popcorn were constructed with genotype data for 503 EUR and 504 EAS individuals sequenced as part of the 1000 Genomes Project. SNPs used had a MAF of at least 5% in both populations, thus resulting in a final set of 3,625,602 SNPs for the 23andMe GWAS sample and 3,642,928 SNPs for the CREAM-EUR sample. These findings support a largely common genetic predisposition to refractive error and myopia in Europeans and Asians, although ancestry-specific risk alleles may exist.

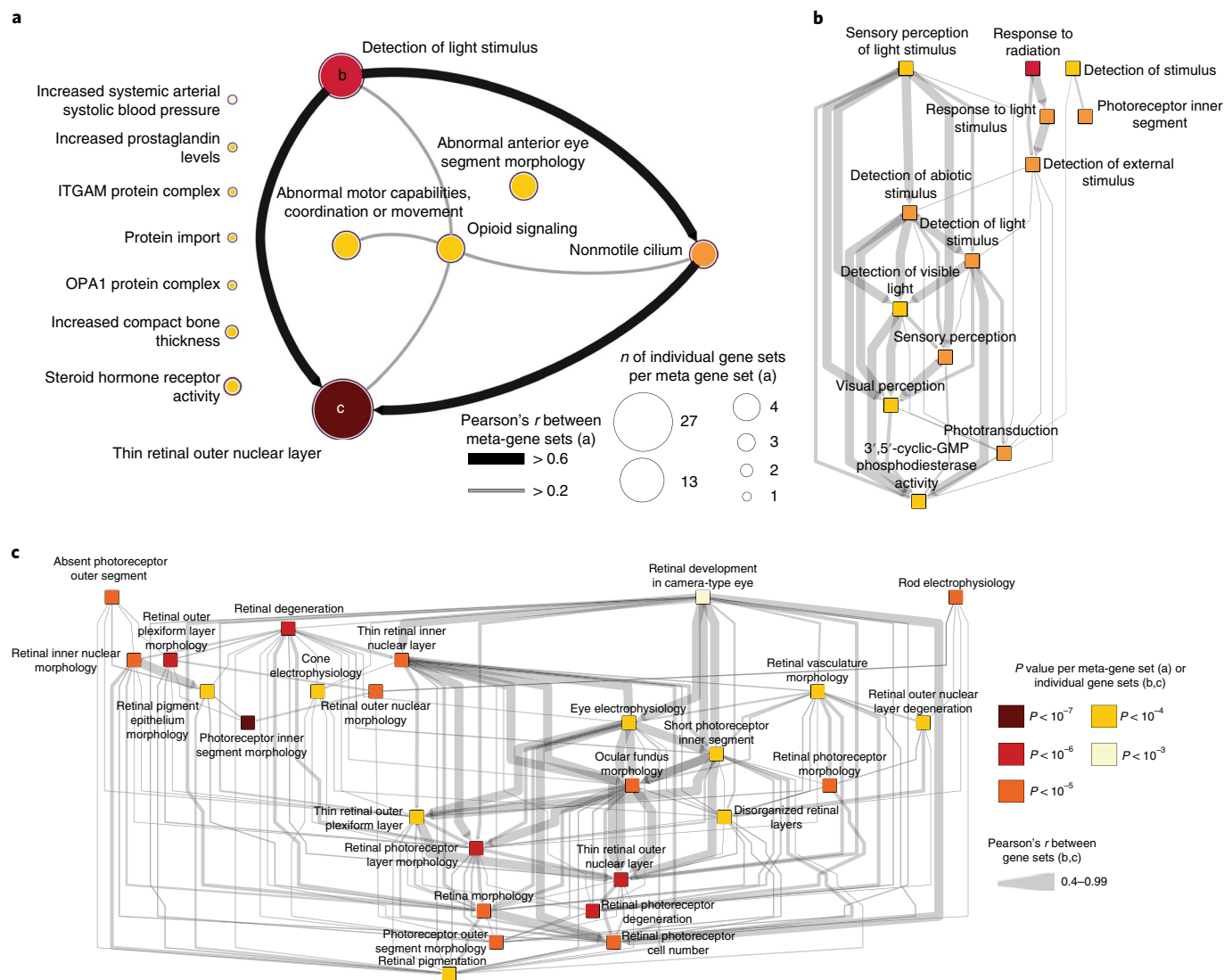


Fig. 4 | Visualization of the DEPICT gene set enrichment analysis based on loci associated with refractive error and the correlation between the (meta) gene sets. a, The 66 significantly enriched reconstituted gene sets clustered into 13 meta-gene sets on the basis of the gene set enrichment analysis of DEPICT (pairwise Pearson correlations; $P < 0.05$). All genetic variants with $P < 1 \times 10^{-5}$ in the GWAS meta-analysis of stage 3 ($n = 21,073$) and an FDR < 0.05 were considered. **b**, Visualization of the interconnectivity among gene sets ($n = 13$; pairwise Pearson correlations; $P < 0.05$) of the meta-gene set 'detection of light stimulus' (GO 0009583). **c**, Visualization of the interconnectivity between gene sets ($n = 27$; pairwise Pearson correlations; $P < 0.05$) of the largest meta-gene set 'thin retinal outer nuclear layer' (MP 0008515). In all panels, (meta)gene sets are represented by nodes colored according to statistical significance, and similarities among them are indicated by edges scaled according to their correlation; Pearson's $r \geq 0.2$ are shown in **a**, and Pearson's $r \geq 0.4$ are shown in **b,c**.

anterior-eye-segment morphology' (MP 0005193; 4 (6%) gene sets). The first three meta-gene sets had a Pearson's correlation > 0.6 . Interestingly, *RGR*, *RP1L1*, *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 ($P = 1.11 \times 10^{-4}$, FDR < 0.01). From the gene prioritization according

Locus	Locus name	Gene priority score	Annotation*				Expression			Biology		Pathways			Known drug target	
			Σ	1	1	1	1	1	1	1	1	1	1	1		
			Internal replication (≥2 cohorts)	Exonic: protein altering	Exonic: non-protein altering	5' or 3' UTR	RNA (nc, sno, linc, other)	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	
<i>GNB3</i>	<i>GNB3</i>	8														
<i>RDH5</i>	<i>BLOC1S1–RDH5, RDH5</i>	7														
<i>CYP26A1</i>	<i>CYP26A1, MYOF</i>	7														
<i>EFEMP1</i>	<i>EFEMP1, PNPT1</i>	7														
<i>GRIA4</i>	<i>GRIA4</i>	7														
<i>RGR</i>	<i>RGR</i>	7														
<i>RORB</i>	<i>RORB</i>	7														
<i>TJP2</i>	<i>TJP2</i>	6														
<i>PRSS56</i>	<i>PRSS56</i>	6														
<i>CABP4</i>	<i>CABP4</i>	6														
<i>FBN1</i>	<i>FBN1</i>	6														
<i>GJD2</i>	<i>GJD2, GOLGA8B</i>	6														
<i>KCNJ2</i>	<i>BC039327, KCNJ2</i>	6														
<i>KCNMA1</i>	<i>KCNMA1</i>	6														
<i>MAF</i>	<i>DYNLRB2, MAF</i>	6														
<i>RCBTB1</i>	<i>RCBTB1</i>	6														
<i>ST18</i>	<i>FAM150A, ST18</i>	6														
<i>TCF7L2</i>	<i>TCF7L2</i>	6														
<i>ZEB2</i>	<i>ZEB2</i>	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 $P < 7.62 \times 10^{-6}$ 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- κ B as the most significant pathway after correction for multiple testing (ratio of the number of molecules, 8.8%; Fisher's exact $P = 1.56 \times 10^{-4}$; 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 $P < 5.4 \times 10^{-4}$ 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-nucleotide-binding protein expressed in rod and cone photoreceptors and ON bipolar cells⁵⁰. *GNB3* participates in signal transduction through G-protein-coupled receptors and enhances the temporal accuracy

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 traits⁷¹ (Supplementary Fig. 12). Refractive error and hyperopia were replicated significantly after correction for multiple testing (adjusted P value = 1.44×10^{-52} and 9.34×10^{-9} , respectively). We found significant overlap with 74 other traits, of which height (adjusted P value = 1.11×10^{-10}), obesity (adjusted P = 1.38×10^{-10}) and body mass index (adjusted P = 4.05×10^{-7}) were most important. Ocular diseases significantly associated were glaucoma (optic cup area and intraocular pressure, adjusted P = 2.69×10^{-5} and 3.01×10^{-5} , respectively) and age-related macular degeneration (adjusted P = 1.27×10^{-3}).

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^{11,72}. 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-analysis⁷³. 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^{11,28,74,75}. 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 group⁷⁶.

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 cells⁷⁷. 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^{65,67,78–87}. *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 effect⁸⁸. 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 cells⁹⁰. 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^{91–93}. 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 ^{65,87,94,95}). 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/ldsc/>; Popcorn, <https://github.com/brielin/Popcorn/>; Online Mendelian Inheritance in Man (OMIM), <http://omim.org/>; wANNOVAR, <http://wannovar.wglab.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

- Pan, C. W., Ramamurthy, D. & Saw, S. M. Worldwide prevalence and risk factors for myopia. *Ophthalmic Physiol. Opt.* **32**, 3–16 (2012).
- Morgan, I. G. What public policies should be developed to deal with the epidemic of myopia? *Optom. Vis. Sci.* **93**, 1058–1060 (2016).
- Morgan, I. & Rose, K. How genetic is school myopia? *Prog. Retin. Eye Res.* **24**, 1–38 (2005).
- Morgan, I. G., Ohno-Matsui, K. & Saw, S. M. Myopia. *Lancet* **379**, 1739–1748 (2012).
- Williams, K. M. et al. Increasing prevalence of myopia in Europe and the impact of education. *Ophthalmology* **122**, 1489–1497 (2015).
- Williams, K. M. et al. Prevalence of refractive error in Europe: the European Eye Epidemiology (E(3)) Consortium. *Eur. J. Epidemiol.* **30**, 305–315 (2015).
- Vongphanit, J., Mitchell, P. & Wang, J. J. Prevalence and progression of myopic retinopathy in an older population. *Ophthalmology* **109**, 704–711 (2002).
- Seet, B. et al. Myopia in Singapore: taking a public health approach. *Br. J. Ophthalmol.* **85**, 521–526 (2001).
- 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).
- Verhoeven, V. J. et al. Visual consequences of refractive errors in the general population. *Ophthalmology* **122**, 101–109 (2015).
- 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).
- Flitcroft, D. I. The complex interactions of retinal, optical and environmental factors in myopia aetiology. *Prog. Retin. Eye Res.* **31**, 622–660 (2012).
- Nakanishi, H. et al. A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1. *PLoS Genet.* **5**, e1000660 (2009).
- 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).
- Stambolian, D. et al. Meta-analysis of genome-wide association studies in five cohorts reveals common variants in RFXO1, a regulator of tissue-specific splicing, associated with refractive error. *Hum. Mol. Genet.* **22**, 2754–2764 (2013).
- Fan, Q. et al. Genetic variants on chromosome 1q41 influence ocular axial length and high myopia. *PLoS Genet.* **8**, e1002753 (2012).
- 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).
- Cheng, C. Y. 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**, 264–277 (2013).
- Shi, Y. et al. Exome sequencing identifies ZNF644 mutations in high myopia. *PLoS Genet.* **7**, e1002084 (2011).
- 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).
- 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).
- Li, Z. et al. A genome-wide association study reveals association between common variants in an intergenic region of 4q25 and high-grade myopia in the Chinese Han population. *Hum. Mol. Genet.* **20**, 2861–2868 (2011).
- 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).
- Tran-Viet, K. N. et al. Mutations in SC02 are associated with autosomal-dominant high-grade myopia. *Am. J. Hum. Genet.* **92**, 820–826 (2013).
- Aldahmesh, M. A. et al. Mutations in LRPAP1 are associated with severe myopia in humans. *Am. J. Hum. Genet.* **93**, 313–320 (2013).
- Verhoeven, V. J. et al. Genome-wide meta-analyses of multiethnic cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat. Genet.* **45**, 314–318 (2013).
- Kiefer, A. K. et al. Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS Genet.* **9**, e1003299 (2013).
- Wojciechowski, R. & Hysi, P. G. Focusing in on the complex genetics of myopia. *PLoS Genet.* **9**, e1003442 (2013).
- 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
- Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
- 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.biorxiv.org/content/early/2017/12/29/240283/> (2017).
- Hsu, F. et al. The UCSC Known Genes. *Bioinformatics* **22**, 1036–1046 (2006).
- Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
- Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
- Kelly, M. P. Does phosphodiesterase 11A (PDE11A) hold promise as a future therapeutic target? *Curr. Pharm. Des.* **21**, 389–416 (2015).
- 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).
- 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).
- 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).
- Bakshi, A. et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci. Rep.* **6**, 32894 (2016).
- 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).
- 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).
- Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- 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, B. C., 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. Ritchey, 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 GNB3 cause a unique form of autosomal-recessive congenital stationary night blindness. *Am. J. Hum. Genet.* **98**, 1011–1019 (2016).
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. Nikonov, S. S. et al. Cones respond to light in the absence of transducin β subunit. *J. Neurosci.* **33**, 5182–5194 (2013).
54. Stone, E. M. et al. A single EFEMP1 mutation associated with both Malattia Leventinese and Doyme honeycomb retinal dystrophy. *Nat. Genet.* **22**, 199–202 (1999).
55. Mackay, D. S., Bennett, T. M. & Shiels, A. Exome sequencing identifies a missense variant in EFEMP1 co-segregating in a family with autosomal dominant primary open-angle glaucoma. *PLoS One* **10**, e0132529 (2015).
56. Springelkamp, H. et al. ARHGGEF12 influences the risk of glaucoma by increasing intraocular pressure. *Hum. Mol. Genet.* **24**, 2689–2699 (2015).
57. Haeseleer, F. et al. Essential role of Ca^{2+} -binding protein 4, a Cav1.4 channel regulator, in photoreceptor synaptic function. *Nat. Neurosci.* **7**, 1079–1087 (2004).
58. Littink, K. W. et al. A novel homozygous nonsense mutation in CABP4 causes congenital cone-rod synaptic disorder. *Invest. Ophthalmol. Vis. Sci.* **50**, 2344–2350 (2009).
59. Grimes, W. N., Li, W., Chávez, A. E. & Diamond, J. S. BK channels modulate pre- and postsynaptic signaling at reciprocal synapses in retina. *Nat. Neurosci.* **12**, 585–592 (2009).
60. Keckeis, S., Reichhart, N., Roubeix, C. & Strauß, O. Anoctamin2 (TMEM16B) forms the Ca^{2+} -activated Cl^- channel in the retinal pigment epithelium. *Exp. Eye Res.* **154**, 139–150 (2017).
61. Prasanna, G., Narayan, S., Krishnamoorthy, R. R. & Yorio, T. Eyeing endothelins: a cellular perspective. *Mol. Cell. Biochem.* **253**, 71–88 (2003).
62. Yamashita, T. et al. Essential and synergistic roles of RP1 and RP1L1 in rod photoreceptor axoneme and retinitis pigmentosa. *J. Neurosci.* **29**, 9748–9760 (2009).
63. Davidson, A. E. et al. RP1L1 variants are associated with a spectrum of inherited retinal diseases including retinitis pigmentosa and occult macular dystrophy. *Hum. Mutat.* **34**, 506–514 (2013).
64. Hawthorne, F. et al. Association mapping of the high-grade myopia MYP3 locus reveals novel candidates UHRF1BP1L, PTPRR, and PPFIA2. *Invest. Ophthalmol. Vis. Sci.* **54**, 2076–2086 (2013).
65. Feldkaemper, M. & Schaeffel, F. An updated view on the role of dopamine in myopia. *Exp. Eye Res.* **114**, 106–119 (2013).
66. Paul, M. L., Graybiel, A. M., David, J. C. & Robertson, H. A. D1-like and D2-like dopamine receptors synergistically activate rotation and *c-fos* expression in the dopamine-depleted striatum in a rat model of Parkinson's disease. *J. Neurosci.* **12**, 3729–3742 (1992).
67. Stone, R. A., Lin, T., Laties, A. M. & Iuvone, P. M. Retinal dopamine and form-deprivation myopia. *Proc. Natl. Acad. Sci. USA* **86**, 704–706 (1989).
68. Gardner, M., Bertranpetit, J. & Comas, D. Worldwide genetic variation in dopamine and serotonin pathway genes: implications for association studies. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **147B**, 1070–1075 (2008).
69. D'Souza, U. M. & Craig, I. W. Functional polymorphisms in dopamine and serotonin pathway genes. *Hum. Mutat.* **27**, 1–13 (2006).
70. Beaulieu, J. M. & Gainetdinov, R. R. The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacol. Rev.* **63**, 182–217 (2011).
71. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
72. Holden, B. A. et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology* **123**, 1036–1042 (2016).
73. Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).
74. Chua, S. Y. et al. Age of onset of myopia predicts risk of high myopia in later childhood in myopic Singapore children. *Ophthalmic Physiol. Opt.* **36**, 388–394 (2016).
75. Williams, K. M. et al. Age of myopia onset in a British population-based twin cohort. *Ophthalmic Physiol. Opt.* **33**, 339–345 (2013).
76. Dolgin, E. The myopia boom. *Nature* **519**, 276–278 (2015).
77. Connaughton, V. Glutamate and glutamate receptors in the vertebrate retina. In: H. Kolb et al. eds. *Webvision: The Organization of the Retina and Visual System* (Webvision, Salt Lake City, UT, USA, 1995).
78. Hung, G. K., Mahadas, K. & Mohammad, F. Eye growth and myopia development: unifying theory and Matlab model. *Comput. Biol. Med.* **70**, 106–118 (2016).
79. Norton, T. T. What do animal studies tell us about the mechanism of myopia-protection by light? *Optom. Vis. Sci.* **93**, 1049–1051 (2016).
80. Weiss, S. & Schaeffel, F. Diurnal growth rhythms in the chicken eye: relation to myopia development and retinal dopamine levels. *J. Comp. Physiol. A* **172**, 263–270 (1993).
81. Stone, R. A., Lin, T., Iuvone, P. M. & Laties, A. M. Postnatal control of ocular growth: dopaminergic mechanisms. *Ciba Found. Symp.* **155**, 45–62 (1990).
82. Morgan, I. G. The biological basis of myopic refractive error. *Clin. Exp. Optom.* **86**, 276–288 (2003).
83. Li, X. X., Schaeffel, F., Kohler, K. & Zrenner, E. Dose-dependent effects of 6-hydroxy dopamine on deprivation myopia, electroretinograms, and dopaminergic amacrine cells in chickens. *Vis. Neurosci.* **9**, 483–492 (1992).
84. Iuvone, P. M., Tigges, M., Stone, R. A., Lambert, S. & Laties, A. M. Effects of apomorphine, a dopamine receptor agonist, on ocular refraction and axial elongation in a primate model of myopia. *Invest. Ophthalmol. Vis. Sci.* **32**, 1674–1677 (1991).
85. 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).
86. Ashby, R. Animal studies and the mechanism of myopia-protection by light? *Optom. Vis. Sci.* **93**, 1052–1054 (2016).
87. Rymer, J. & Wildsoet, C. F. The role of the retinal pigment epithelium in eye growth regulation and myopia: a review. *Vis. Neurosci.* **22**, 251–261 (2005).
88. Chen, 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).
89. Chen, P. S. et al. Effects of C825T polymorphism of the GNB3 gene on availability of dopamine transporter in healthy volunteers: a SPECT study. *Neuroimage* **56**, 1526–1530 (2011).
90. Scott, M. S. & Ono, M. From snoRNA to miRNA: dual function regulatory non-coding RNAs. *Biochimie* **93**, 1987–1992 (2011).
91. McFadden, S. A. Understanding and treating myopia: what more we need to know and future research priorities. *Optom. Vis. Sci.* **93**, 1061–1063 (2016).
92. Smith, E. L. III, Hung, L. F. & Arumugam, B. Visual regulation of refractive development: insights from animal studies. *Eye (Lond.)* **28**, 180–188 (2014).
93. Zhang, Y. & Wildsoet, C. F. RPE and choroid mechanisms underlying ocular growth and myopia. *Prog. Mol. Biol. Transl. Sci.* **134**, 221–240 (2015).
94. 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).
95. 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 UKEV for their generosity in sharing data and help in the production of this publication. Funding for this particular GWAS mega-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.C.W.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.S., E.M.v.L., J.E.B.-W., J.Y.T., N.A.F., Q.F., S.-M.S. and V.V. critically reviewed the manuscript. A.N., A.P.K., A.T., C.B., C. Gieger, C.L.S., C.-Y.C., G. Biino, G.C.-P., I.R., J.E.B.W., J.E.H., J. S. Ried, J.W., J.X., K.M.W., K.Y., P.M.C., S.M.H., M.S.T., N.A.F., N.E., P.C., P. Gharakhani, P.K.J., Q.F., R. Höhn, R.L.S., R.P.I., R.W., T.H., T.-H.S.-A., T.Z., 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. Grazal, C.H., C.J.H., C.W., C.-Y.C., D.A.M., F.R., G. Bencic, H.M.-H., J.A.G., J.B.J., J.E.B.-W., J.E.C., J.F.W., J.H.L., J.R.V., J. S. Rahi, J. S. Ried, J.Y.T., K.Y., M.A.M.-S.,

N.G.M., N.P., O. Polašek, O. Pärssinen, O.T.R., P. Gupta, P.J.F., 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., P.J.v.d.S. and R. Hask performed *in silico* pathway analyses. C.C.W.K. and C.J.H. conceived and designed the outline of the current report, and supervised conduction of experiments and analyses jointly with A.M., A.H., A.W.H., C.D., C.H., C.J.H., C.M.v.D., C.W., C.-Y.C., D.A.M., D.S., E.-S.T., F.M., G. Biino, I.R., J.A.G., J.B.J., J.E.B.-W., J.E.C., J.F.W., J.H.L., J.R.V., J.Y.T., N.A., N.A.F., N.P., O. Pärssinen, O.T.R., P.J.F., P.N.B., S.K.I., S.-M.S., T.L., T.Y.W., T.L.Y., V.V., Y.X.W. and Y.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.J. is a patent holder with

Biocompatibles UK Ltd. (Franham, 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 included in 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.

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

¹Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands. ²Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. ³Department of Epidemiology and Medicine, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁴Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ⁵Wilmer Eye Institute, Johns Hopkins Medical Institutions, Baltimore, MD, USA. ⁶Section of Academic Ophthalmology, School of Life Course Sciences, King's College London, London, UK. ⁷23andMe, Inc., Mountain View, CA, USA. ⁸Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands. ⁹Department of Ophthalmology and Visual Sciences, University of Wisconsin–Madison, Madison, WI, USA. ¹⁰Centre for Quantitative Medicine, DUKE–National University of Singapore, Singapore, Singapore. ¹¹Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ¹²NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK. ¹³Ocular Epidemiology Research Group, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. ¹⁴Department of Ophthalmology, University Hospital Bern, Inselspital, University of Bern, Bern, Switzerland. ¹⁵Department of Ophthalmology, University Medical Center Mainz, Mainz, Germany. ¹⁶Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹⁷Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA. ¹⁸Estonian Genome Center, University of Tartu, Tartu, Estonia. ¹⁹Department of Ophthalmology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ²⁰Department of Public Health, University of Helsinki, Helsinki, Finland. ²¹Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karls-University of Heidelberg, Mannheim, Germany. ²²Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China. ²³Centre for Eye Research Australia, Ophthalmology, Department of Surgery, University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia. ²⁴Department of Ophthalmology, Centre for Vision Research, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. ²⁵Program in Genetics and Genome Biology, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada. ²⁶School of Optometry & Vision Sciences, Cardiff University, Cardiff, UK. ²⁷Department of Population Health Sciences, Bristol Medical School, Bristol, UK. ²⁸Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore. ²⁹Saw Swee Hock School of Public Health, National University Health Systems, National University of Singapore, Singapore, Singapore. ³⁰Department of Health Service Research, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. ³¹Statistics Support Platform, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. ³²Life Sciences Institute, National University of Singapore, Singapore, Singapore. ³³MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK. ³⁴Faculty of Medicine, University of Split, Split, Croatia. ³⁵Department of Ophthalmology, Sisters of Mercy University Hospital, Zagreb, Croatia. ³⁶Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ³⁷A list of members and affiliations appears at the end of the paper. ³⁸Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. ³⁹Clinic for General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany. ⁴⁰Department of Bioinformatics, Erasmus Medical Center, Rotterdam, The Netherlands. ⁴¹Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands. ⁴²Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands. ⁴³Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA. ⁴⁴Department of Ophthalmology and Visual Sciences, Case Western Reserve University and University Hospitals Eye Institute, Cleveland, OH, USA. ⁴⁵Department of Genetics, Case Western Reserve University, Cleveland, OH, USA. ⁴⁶Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁴⁷Netherlands Consortium for Healthy Ageing, Netherlands Genomics Initiative, The Hague, The Netherlands. ⁴⁸Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. ⁴⁹Department of Clinical Chemistry, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. ⁵⁰Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere, Tampere, Finland. ⁵¹Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. ⁵²Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. ⁵³Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. ⁵⁴Institute for Maternal and Child Health–IRCCS ‘Burlo Garofolo’, Trieste, Italy. ⁵⁵Department of Medical and Molecular Genetics, Indiana University, School of Medicine, Indianapolis, IN, USA. ⁵⁶Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ⁵⁷Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ⁵⁸Department of Ophthalmology, Flinders University, Adelaide, South Australia, Australia. ⁵⁹Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. ⁶⁰Great Ormond Street Institute of Child Health, University College London, London, UK. ⁶¹Ulverschroft Vision Research Group, University College London, London, UK. ⁶²Université de Bordeaux, Inserm, Bordeaux Population Health Research Center, team LEHA, UMR 1219, F-33000 Bordeaux, France. ⁶³Institut Pasteur de Lille, Lille, France. ⁶⁴Inserm, U1167, RID–AGE–Risk factors and molecular determinants of aging–related diseases, Lille, France. ⁶⁵Université de Lille, U1167–Excellence Laboratory LabEx DISTALZ, Lille, France. ⁶⁶Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. ⁶⁷Department of Ophthalmology, Academic Medical Center, Amsterdam, The Netherlands. ⁶⁸Netherlands Institute for Neurosciences (NIN–KNAW), Amsterdam, The Netherlands. ⁶⁹Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. ⁷⁰Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. ⁷¹Academic Medicine Research Institute, Singapore, Singapore. ⁷²Retino Center, Singapore National Eye Centre, Singapore, Singapore. ⁷³Department of Ophthalmology, Menzies Institute of Medical Research, University of Tasmania, Hobart, Tasmania, Australia. ⁷⁴Centre for

Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Western Australia, Australia. ⁷⁵Department of Genetics, Genomics and Informatics, University of Tennessee Health Sciences Center, Memphis, TN, USA. ⁷⁶Department of Ophthalmology, Central Hospital of Central Finland, Jyväskylä, Finland. ⁷⁷Gerontology Research Center, Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland. ⁷⁸Myopia Research Group, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. ⁷⁹Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands. ⁸⁰These authors contributed equally: Milly S. Tedja, Robert Wojciechowski, Pirro G. Hysi, Nicholas Eriksson, Nicholas A. Furlotte, Virginie J. M. Verhoeven. ⁸¹These authors jointly supervised this work: Jeremy A. Guggenheim, Joyce Y. Tung, Christopher J. Hammond, Caroline C. W. Klaver. *e-mail: c.c.w.klaver@erasmusmc.nl

The CREAM Consortium

Tin Aung^{82,83}, Amutha B. Veluchamy^{82,84}, Kathryn P. Burdon⁵⁸, Harry Campbell³⁶, Li Jia Chen⁸⁵, Peng Chen⁸³, Wei Chen⁸⁶, Emily Chew⁴⁵, Margaret M. Deangelis⁸⁷, Xiaohu Ding⁸⁸, Angela Döring⁶⁶, David M. Evans^{89,90}, Sheng Feng⁹¹, Brian Fleck⁹², Rhys D. Fogarty⁵⁸, Jeremy R. Fondran⁴³, Maurizio Fossarello⁹³, Xiaobo Guo^{88,94}, Annet E. G. Haarman^{1,2}, Mingguang He^{23,88}, Laura D. Howe^{90,95}, Sarayut Janmahasatian⁴³, Vishal Jhanji⁸⁵, Mika Kähönen⁹⁶, Jaakko Kaprio^{20,97}, John P. Kemp⁹⁰, Kay-Tee Khaw¹¹, Chiea-Chuen Khor^{29,83,87,98}, Eva Krapohl⁹⁹, Jean-François Korobelnik^{100,101}, Kris 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, Songhomitra Panda-Jonas¹⁰³, Chi Pui Pang⁸⁵, Mario Pirastu¹⁰⁵, Robert Plomin⁹⁹, Taina Rantanen⁷⁷, Maria Schache²³, Ilkka Seppälä⁴⁹, George D. Smith⁹⁰, Beate St Pourcain^{90,106}, Pancy O. Tam⁸⁵, J. Willem L. Tideman^{1,2}, Nicholas J. Timpson⁹⁰, Simona Vaccargiu¹⁰⁵, Zoran Vatavuk³⁵, Jie Jin Wang^{23,24}, Ningli Wang²², Nick J. Wareham¹⁰⁷, Alan F. Wright^{4,33}, Liang Xu²², Maurice K. H. Yap¹⁰⁸, Seyhan Yazar⁷⁴, Shea 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 Vacic⁷ 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 Chakravarthy¹¹⁵, 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 Petzold¹³⁰, Alicja R. Rudnicka¹²⁹, Jay E. Self^{122,131}, 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

⁸²Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. ⁸³Department of Ophthalmology, National University Health Systems, National University of Singapore, Singapore, Singapore. ⁸⁴Duke-NUS Medical School, Singapore, Singapore, Singapore. ⁸⁵Department of Ophthalmology and Visual Sciences, Chinese University of Hong Kong, Hong Kong Eye Hospital, Kowloon, Hong Kong. ⁸⁶School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, China. ⁸⁷Department of Ophthalmology and Visual Sciences, John Moran Eye Center, University of Utah, Salt Lake City, UT, USA. ⁸⁸State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. ⁸⁹Translational Research Institute, University of Queensland Diamantina Institute, Brisbane, Queensland, Australia. ⁹⁰MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. ⁹¹Department of Pediatric Ophthalmology, Duke Eye Center For Human Genetics, Durham, NC, USA. ⁹²Princess Alexandra Eye Pavilion, Edinburgh, UK. ⁹³University Hospital 'San Giovanni di Dio', Cagliari, Italy. ⁹⁴Department of Statistical Science, School of Mathematics, Sun Yat-Sen University, Guangzhou, China. ⁹⁵School of Social and Community Medicine, University of Bristol, Bristol, UK. ⁹⁶Department of Clinical Physiology, Tampere University Hospital and School of Medicine, University of Tampere, Tampere, Finland. ⁹⁷Institute for Molecular Medicine Finland FIMM, HiLIFE Unit, University of Helsinki, Helsinki, Finland. ⁹⁸Division of Human Genetics, Genome Institute of Singapore, Singapore, Singapore. ⁹⁹MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ¹⁰⁰Université de Bordeaux, Bordeaux, France. ¹⁰¹Institut National de la Santé Et de la Recherche Médicale (INSERM), Institut de Santé Publique d'Épidémiologie et de Développement (ISPED), Centre INSERM U897-Epidémiologie-Biostatistique, Bordeaux, France. ¹⁰²Department of Ophthalmology, Yokohama City University School of Medicine, Yokohama, Japan. ¹⁰³Suraj Eye Institute, Nagpur, Maharashtra, India. ¹⁰⁴Institute of Neurogenomics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany. ¹⁰⁵Institute of Genetic and Biomedic Research, National Research Council, Cagliari, Italy. ¹⁰⁶Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. ¹⁰⁷MRC Epidemiology Unit, Institute of Metabolic Sciences, University of Cambridge, Cambridge, UK. ¹⁰⁸Centre for Myopia Research, School of Optometry, Hong Kong Polytechnic University, Hong Kong, Hong Kong. ¹⁰⁹Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong, Hong Kong. ¹¹⁰Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. ¹¹¹School of Computer Science and Mathematics, Kingston University, Surrey, UK. ¹¹²Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK. ¹¹³Primary Care & Public Health Sciences, King's College London, London, UK. ¹¹⁴Faculty of Medicine University of Southampton, Southampton General Hospital, Southampton, UK. ¹¹⁵School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, Northern Ireland, UK. ¹¹⁶Optometry and Visual Science, School of Health Science, City, University of London, London, UK. ¹¹⁷Division of Health Sciences & Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. ¹¹⁸School of Clinical Sciences, Faculty of Medicine and Dentistry, University of Bristol, Bristol, UK. ¹¹⁹Department of Psychiatry, Oxford University, Warneford Hospital, Oxford, UK. ¹²⁰Department of Eye and Vision Science, University of Liverpool, Liverpool, UK. ¹²¹Centre for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, UK. ¹²²Department of Ophthalmology, University of Southampton NHS Foundation Trust, Southampton, UK. ¹²³Edinburgh Imaging, University of Edinburgh, Edinburgh, UK. ¹²⁴Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK. ¹²⁵Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. ¹²⁶Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, UK. ¹²⁷Department of Ophthalmology, Leeds Teaching Hospitals NHS Trust, Leeds, UK. ¹²⁸Department of Ophthalmology, King's College Hospital NHS Foundation Trust, London, UK. ¹²⁹St George's, University of London, London, UK. ¹³⁰UCL Institute of Neurology, London, UK. ¹³¹Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. ¹³²Institute of Genetic Medicine, Newcastle University, Newcastle Upon Tyne, UK. ¹³³Gloucestershire Retinal Research Group, Gloucestershire Hospitals NHS Foundation Trust, Cheltenham General Hospital, Cheltenham, UK. ¹³⁴Centre for Medical Informatics, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ¹³⁵School of Science and Engineering, University of Dundee, Dundee, UK. ¹³⁶Nottingham University Hospitals NHS Trust, Nottingham, UK. ¹³⁷Norwich Medical School, University of East Anglia, Norwich, Norfolk, UK.

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's 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 = spherical refractive error +1/2 cylinder refractive error) was the outcome variable for CREAM; myopic refractive error was represented by self-reported AODM for 23andMe²⁷.

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 minimac³⁶ or IMPUTE³⁷. 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 analyses^{26,27}; 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 relatedness³⁸. For the 23andMe participants, Cox proportional hazards analysis testing with AODM as the dependent variable was performed as previously described²⁷, and P was calculated with a likelihood-ratio test for the single-marker genotype term. We used an additive SNP allelic-effect model for all analyses.

Centralized quality control per study. After individual GWAS, all studies were subjected to a second round of QC. Quantile–quantile, effect-allele frequency, P - Z test, standard error–sample size, and genomic-control inflation-factor plots were generated for each individual cohort in EasyQC³⁹ (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 $MAF < 1\%$ or imputation quality score < 0.3 (info metric of IMPUTE) or $Rsq < 0.3$ (minimac) were excluded. A fixed-effects inverse-variance-weighted meta-analysis was performed in METAL¹⁰⁰. 1,063 variants clustering in 24 loci (Supplementary Table 2) were genome-wide significant ($P = 5.0 \times 10^{-8}$). All 37 loci that were previously found by CREAM and 23andMe by using genotype data imputed to the HapMap II reference panel were replicated ($P_{\text{Bonferroni}} 1.85 \times 10^{-3}$), and 36 of the 37 were genome-wide significant^{26,27} (Supplementary Table 2). In stage 2, a meta-analysis of the two 23andMe cohorts ($n_{23andMe_V2} = 12,128$; $n_{23andMe_V3} = 92,165$) 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 (at $P < 0.05$), we also performed a random-effects meta-analysis in METASOFT⁴⁹. 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. 2b), and thus weighting by $(1/\sqrt{n_{\text{effective}}})$ yielded a final weighting ratio of 1:1 (ref. 101). Genome-wide statistical significance was defined at $P < 5.0 \times 10^{-8}$ (ref. 102).

All three meta-analysis stages were performed under genomic control. Study-specific and meta-analysis lambda (λ) 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 constructed³⁰ (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 'qqman'), regional, box and forest plots were made in R version 3.2.3 and LocusZoom¹⁰³. An overview of the Hardy–Weinberg P of all index variants per cohort can be found in Supplementary Table 4. The comparison between refractive error and age of onset was performed in the LDSC program³⁰.

Population stratification and heritability calculations. Each study assessed the degree of genetic admixture and stratification in study participants through the use of principal components. Homogeneity of participants was ensured by removal of all individuals whose ancestry did not match the prevailing ancestral group. We used genomic inflation factors to control for admixture and stratification, and performed genomic-controlled meta-analysis to account for the effects of any residual heterogeneity. To further distinguish between inflation from a true polygenic signal and population stratification, we examined the relationship between test statistics and LD with LDSC. CREAM-EUR, CREAM-ASN and 23andMe were evaluated separately; variants not present in HapMap3 and with $MAF < 1\%$ were excluded. SNP heritability estimates were calculated in LDSC for the same set of genetic variants.

Locus definition and annotation. All study effect-size estimates were oriented to the positive strand of the NCBI Build 37 reference sequence of the human genome. The index variant of a locus was defined as the variant with the lowest P in a 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 ANNOVAR¹⁰⁴ based on UCSC Known Gene Database³⁴. 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 (PhyloP¹⁰⁵) and prediction tools (SIFT³⁸, MutationTaster¹⁰⁶, align GVGD^{39,40} 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, by using GCTA-COJO³². We transformed the Z scores of the summary statistics to betas with the following formula: standard error = $\sqrt{1/2N \times MAF(1-MAF)}$. We performed the GCTA-COJO analysis³² by using summary-level statistics from the meta-analysis on all cohorts. LD between variants was estimated from RSI-III.

Replication in UK Biobank. The UKEV 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 disorders³³. 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 used¹⁰⁷, including age, sex, genotyping array and the first ten principal components as covariates. 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 UKEV, 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 not available for replication: one variant (rs188159083) was neither present on the array nor was a surrogate available in UKEV, and five variants showed evidence of departure from HWE (HWE exact test $P < 3.0 \times 10^{-4}$).

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 fastBAT⁴¹ 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 $P < 2 \times 10^{-6}$ (Bonferroni correction based on 25,000 genes) and per-variant $P > 5 \times 10^{-8}$. Second, we applied another gene-based test in EUGENE⁴², which includes only variants that are eQTLs (Genotype

Tissue Expression (GTEx) data, blood¹⁰⁹). 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 $P < 2 \times 10^{-6}$ (and not found in the single variant analysis) were considered significant. Finally, we used functional annotation information from genome-wide-significant loci to reweight results in fgwas (version 0.3.64 (ref. 43)). Fgwas incorporates functional annotation (for example, DNase I-hypersensitive sites in various tissues and 3'-UTR regions) to reweight 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 PLINK¹¹⁰, 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, RSIII and RSIII ($n = 10,792$), PGRS values were calculated with the --score command in PLINK across the following strata of P thresholds: 5.0×10^{-8} , 5.0×10^{-7} , 5.0×10^{-6} , 5.0×10^{-5} , 5.0×10^{-4} , 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 calculated for myopia (SphE ≤ -3 s.d.) vs. hyperopia (SphE $\geq +3$ s.d.).

Genetic correlation between ancestries. We used Popcorn⁴⁷ 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 PLINK¹¹⁰ to remove indels, strand-ambiguous variants, variants without an 'rs' ID prefix and variants located in the MHC region on chromosome 6 (chromosome 6: 25000000–33500000; 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 diopters 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; $P_{\text{Bonferroni}} 1.19 \times 10^{-4}$); evidence of eQTL from FUMA³² 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, DisGeNET¹¹¹); 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 (DEPICT⁴⁸); presence of the gene in the gene priority 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 (SuperTarget¹¹², STITCH¹¹³, DrugBank¹¹⁴ and PharmaGkb¹¹⁵). 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 ≥ 2 .

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.w.klaver@erasmusmc.nl (CREAM) and/or apply.research@23andMe.com (23andMe) for more information and to access the data.

References

- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
- 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).
- Chen, W. M. & Abecasis, G. R. Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* **81**, 913–926 (2007).
- Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Zaykin, D. V. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* **24**, 1836–1841 (2011).
- Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**, 227–234 (2008).
- Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
- Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* **10**, 1556–1566 (2015).
- Cooper, G. M. et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
- Schwarz, J. M., Rödelberger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
- Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
- McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
- Consortium, G. T., GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
- 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).
- Günther, S. et al. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* **36**, D919–D922 (2008).
- Kuhn, M. et al. STITCH 4: integration of protein-chemical interactions with user data. *Nucleic Acids Res.* **42**, D401–D407 (2014).
- Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
- Whirl-Carrillo, M. et al. Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **92**, 414–417 (2012).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

▶ 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 2.00×10^{-3}). Vice versa, 29 (25.9%) of the loci identified at Stage 2 replicated in Stage 1 (pBonferroni 4.46×10^{-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-analysis 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 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The <u>exact</u> sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly. |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement indicating how many times each experiment was replicated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> 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) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> 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 (<http://genetics.bwh.harvard.edu/pph2/>); SIFT (http://sift.jcvi.org/www/SIFT_aligned_seqs_submit.html); Mutation Taster (<http://www.mutationtaster.org/>); 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.