Meta-analysis of 542,934 subjects of European ancestry identifies 336 novel genes and mechanisms predisposing to refractive error and myopia

Pirro G. Hysi[1,2,3 †], Hélène Choquet[4†], Anthony P. Khawaja [5,6†], Robert Wojciechowski[7,8†], Milly S Tedja[9,10†], Jie Yin[4], Mark J. Simcoe[2], Karina Patasova[1], Omar A. Mahroo[1,5], Khanh K Thai[4], Phillippa M Cumberland[3,12], Ronald B. Melles[13], Virginie J.M. Verhoeven[9,10,11], Veronique Vitart[14], Ayellet Segre[15], Richard A. Stone[16], Nick Wareham[6], Alex W Hewitt[17], David A Mackey[17,18], Caroline CW Klaver[9,10,19,20], Stuart MacGregor[21], The Consortium for Refractive Error and Myopia, Peng Khaw[5], Paul J. Foster[5,22], The UK Eye and Vision Consortium, Jeremy A. Guggenheim[23], 23andMe Inc., Jugnoo S Rahi[3,5,12,24*], Eric Jorgenson[4*], and Christopher J Hammond[1,2 *]

Author Information:

1 King's College London, Section of Ophthalmology, School of Life Course Sciences, London, UK
2 King's College London, Department of Twin Research and Genetic Epidemiology, London, UK
3 University College London, GOSH Institute of Child Health, London, UK
4 Division of Research, Kaiser Permanente Northern California, Oakland, California, USA
5 NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK
6 Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK
7 Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA
8 Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD, USA
9 Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands
10 Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
11 Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands
12 Ulverscroft Vision Research Group, UCL Great Ormond Street Institute of Child Health, University College London, UK
13 Kaiser Permanente Northern California, Department of Ophthalmology, Redwood City, CA, USA
14 Institute of Genetics and Molecular Medicine, The University of Edinburgh, United Kingdom
15 Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear, Boston, MA, USA
16 University of Pennsylvania School of Medicine, Philadelphia, PA, USA
17 Department of Ophthalmology, Royal Hobart Hospital, Hobart, Tasmania
18 Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, Perth, WA, WA 6009, Australia
19 Dept. Ophthalmology, Radboud University Medical Center, Rotterdam
20 Institute of Molecular and Clinical Ophthalmology Basel, Switzerland
21 QIMR Berghofer Medical Research Institute, Brisbane, Australia
22 Division of Genetics and Epidemiology, UCL Institute of Ophthalmology, London, UK.
23 Cardiff University, School of Optometry & Vision Sciences, UK
24 Department of Ophthalmology and NIHR Biomedical Research Centre Great Ormond Street Hospital NHS Foundation Trust

† *These authors jointly led this work*
* *These authors jointly supervised this work*

Abstract

Refractive errors, in particular myopia, are a leading cause of morbidity and disability world-wide and their prevalence is rising, largely due to cultural and environmental changes. Genetic investigation is a valuable tool to better understand the molecular mechanisms underlying abnormal eye development and impaired vision. We conducted a meta-analysis of genome-wide association studies involving 542,934 European participants and identified 336 novel genetic loci associated with refractive error that explain an additional 4.6% of spherical equivalent heritability, or an improvement by a third over the previous estimates. Collectively, all associated genetic variants explain 18.4% of heritability and improve the accuracy of myopia prediction (AUC=0.75). Our results suggest that refractive error is genetically heterogeneous, driven by genes participating in the development of every anatomical component of the eye. In addition, our analyses suggest that genetic factors controlling circadian rhythm and pigmentation are also involved in the development of myopia and refractive error. These results may make possible predicting refractive error and the development of personalized myopia prevention strategies in the future.

Refractive errors (RE) occur when converging light rays from an image do not clearly focus on the retina. They are the seventh most prevalent clinical condition[1] and the second leading cause of disability in the world[2]. The prevalence of RE is rapidly increasing, mostly driven by a dramatic rise in the prevalence of one of its forms, myopia (near-sightedness). Although the causes of such a rise over a short time are likely due to environmental and cultural changes from the mid-20th century[3], RE are highly heritable[4]. Several studies[5,6] have previously sought to identify genes controlling molecular mechanisms leading to RE and myopia. However, the variance and heritability that can be attributed to known genetic factors is modest[7] and our knowledge of pathogenic mechanisms remains partial. Here, we conduct a meta-analysis combining data from quantitative spherical equivalent and myopia status from large and previously unpublished genome-wide association studies (GWAS) of more than half a million subjects from the UK Biobank, 23andMe and the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohorts, with subsequent replication and meta-analysis with data previously reported from the Consortium for Refractive Error and Myopia (CREAM).

Results

Association Results.

Analyses were restricted to subjects of European ancestry (Supplementary Figure 1) and combined results from quantitative measures of spherical equivalent and categorical myopia status. Spherical equivalent quantifies RE; a negative spherical equivalent, below a certain threshold defines myopia. We used results obtained from GWAS of directly measured spherical equivalent in 102,117 population-based UK Biobank participants[8], and 34,998 subjects participating in the GERA Study[9] and combined them with results of analyses of self-reported myopia in 106,086 cases and 85,757 controls from the customer base of 23andMe, Inc. (Mountain View, CA), a personal genomics company[10]. Additionally, we included results from an analysis on the refractive status inferred using demographic and self-reported information on age at first use of prescription glasses among the UK Biobank participants not contributing to the quantitative GWAS (108,956 likely myopes to 70,941 likely non-myopes, see Supplementary Methods). All analyses were adjusted for age, sex and main principal components. To obtain an overall association with RE, we meta-analyzed the results from all studies by using the z-scores

from the GWAS of the spherical equivalent and the negative values of z-scores from the case-control studies (23andMe and UK Biobank), since myopia is negatively correlated with spherical equivalent. As expected, the large total sample size of the discovery meta-analysis (N=508,855) led to a nominally large genomic inflation factor (λ=1.94). The LD score regression intercept was (1.17), and the (intercept-1)/(mean(chi^2)-1) ratio of 0.097 is fully in line with the expectations of polygenicity[11].

We found associations for 438 discrete genomic regions (Figure 1, Supplementary Table 1), defined by markers contiguously associated at conventional level of GWAS significance[12,13] of $p<5x10^{-08}$, separated by more than 1 Mbp from other GWAS-associated markers, as recommended elsewhere[14]. Among them, 308 loci, including 14 on chromosome X, were not described in previous GWAS studies of refractive error[7]. The observed effect sizes were consistent across all the studies (Supplementary Table 1 and Supplementary File 1). The association with RE was statistically strongest for rs12193446 ($p=9.87x10^{-328}$), within *LAMA2*, a gene previously associated with RE[5,6], mutations of which cause muscular dystrophy[15]. Consistent with these *LAMA2* properties, polymorphisms located within the genes coding for both major *LAMA2* receptors, *DAG1*[16] ($p=1.67x10^{-08}$ for rs111327216) and *ITGA7*[17] ($p=8.57x10^{-09}$ for rs17117860) which are also known causes of muscular dystrophy[18,19], were significantly associated with RE in the discovery meta-analysis.

We compared our discovery meta-analysis findings with GWAS results from 34,079 participants in the CREAM consortium, who were part of a previously reported meta-analysis[7]. To avoid any potential overlap with the UK Biobank participants, only non-UK European CREAM participants were used for replication. Despite the vast power differential, 55 of the SNPs that showed the strongest association in their respective regions in the discovery meta-analysis were significant after Bonferroni correction in the replication sample. A further 142 had a false discovery rate (FDR) < 0.05 and 192 were nominally significant at P < 0.05 (Supplementary Table 2). The effect sizes observed in the discovery and replication samples were strongly correlated (Pearson's r=0.91, Supplementary Figure 2). Meta-analysis of all five cohorts (discovery and replication) expanded the number to 449 associated of regions of variable length and number of SNPs (Supplementary Figure 3), of which 336 regions were novel (Supplementary Table 3).

Most of the 449 RE-associated regions contained at least one gene linked to severe ocular manifestations in the Online Mendelian Inheritance In Man (OMIM) resource or other genes with interesting link to eye disease (Supplementary Table 4). Although most loci identified through our meta-analyses were novel, several of them hosted genes that harbor mutations leading to myopia or other RE phenotypes. Several genes significantly associated with RE were linked to Mendelian disorders affecting corneal structure, some of which code for transcription factors involved in corneal development[20] (Supplementary Table 5). Mutations in these genes cause corneal dystrophies (*SLC4A11*, $p=5.81x10^{-11}$ for rs41281858, *TCF4*, $p=4.14x10^{-08}$, rs41396445; *LCAT*, $p=1.26x10^{-10}$, rs5923; and *DCN*, $p=3.67x10^{-09}$, rs1280632), megalocornea (*LTBP2*, $p=1.91x10^{-24}$, rs73296215) and keratoconus (*FNDC3B*, $p=1.89x10^{-14}$, rs199771582, previously described[7]). Eleven RE-associated genes were linked to anomalies of the crystalline lens (Supplementary Table 6), including genes linked to autosomal dominant cataracts (*PAX6* previously linked to myopia[21], $p=8.31x10^{-11}$, rs1540320; *PITX3*, $p=1.05x10^{-10}$, rs7923183; *MAF*, $p=5.50x10^{-09}$, rs16951312; *CHMP4B*, $p=9.95x10^{-11}$, rs6087538; *TDRD7*, $p=4.79x10^{-08}$, rs13301794) and lens ectopia (*FBN1*, $p=3.30x10^{-24}$, rs2017765; *ADAMTSL4*, $p=8.19x10^{-14}$, rs12131376). Some of the genes affected several eye components. For example, *LTBP2 variants* are also associated with congenital glaucoma[22], and *COL4A3* (rs7569375, $p=1.14x10^{-08}$) causes Alport syndrome, which manifests with abnormal lens shape (lenticonus) and structural changes in the retina.

Association was also observed within or near 13 genes known to harbor mutations causing microphthalmia (Supplementary Table 7), including *TENM3* ($p=2.48x10^{-11}$, rs35446926); *OTX2*

(p=6.15x10^{-11}, rs928109); *VSX2*, (p=4.60x10^{-10}, rs35797567); *MFRP*, (p=2.85x10^{-16}, rs10892353) and the previously identified[6] *TMEM98*, (p=3.49x10^{-43}, rs62067167). Association was also found for *VSX1* (p=4.59x10^{-08} for rs6050351), a gene that is closely regulated by VSX2[23] and believed to play important roles in eye development[24]. Many of the genes nearest associated SNPs have been linked to inherited retinal disease (Supplementary Table 8), including 32 genes linked to cone-rod dystrophies, night blindness and retinitis pigmentosa, and age-related macular degeneration (*HTRA1/ARMS2*). Among genes in novel regions associated with RE, *ABCA4* (p=3.20x10^{-10} for rs11165052), and ARMS2/HTRA1 (p=5.72x10^{-23} for rs2142308) are linked to macular disorders and numerous others to retinitis pigmentosa, retinal dystrophy and other retinal diseases, such as *FBN2*, (p=8.63x10^{-11}, rs6860901) , *TRAF3IP1* (p=5.71x10^{-16}, rs7596847), CWC27 (p=1.84x10^{-18}, rs1309551). Significant association was found near other genes of interest such as *DRD1* (p=4.51x10^{-16}, rs13190379), a dopamine receptor. Together, these results are consistent with previous suggestions of light transmission and transduction in RE[7,25].

Wnt signaling has previously been implicated in experimental myopia[26]. We found significant association near several *Wnt* protein-coding genes (*WNT7B*, a gene previously associated with axial length[27], p=1.42x10^{-26} for rs73175083; *WNT10A*, previously associated with central corneal thickness[28], p=1.65x10^{-17} for rs121908120 and *WNT3B*, p=8.52x10^{-16} for rs70600), suggesting that organogenesis through *Wnt* signaling is likely to be involved in RE. Significant association were found at genes coding for key canonical (e.g. rs13072632 within the *CTNNB1* gene, p=7.30x10^{-27}; *AXIN2*, rs9895291, p=1.40x10^{-08}) and non-canonical *Wnt* pathway members (*NFATC3*, rs147561310, p=1.493x10^{-12}) or at genes coding for both (*RHOA*, rs7623687, p=1.81x10^{-11} or the previously described[7] *TCF7L2*, rs56299331, p=9.38x10^{-46}; Supplementary Table 9).

Similar to previous published analyses[25], we found associations for genes involved in sodium, potassium, calcium magnesium and other cation transporters (Supplementary Table 10). The involvement of genes related to glutamatergic synaptic transmission was also notable (Supplementary Table 11). Glutamate is a first synapse transmitter released by photoreceptors towards bipolar cells and is the main excitatory neurotransmitter of the retina, and expression of genes participating in glutamate signaling pathways is significantly altered in myopia models[29]. These associations support the involvement in RE pathogenesis of neurotransmission and neuronal depolarization and hyperpolarization that was also suggested before[7]. Associations with *POU6F2* gene intronic variants (rs2696187, p=1.11x10^{-11}) also suggests involvement of factors related to development of amacrine and ganglion cells[30]. Other genes at RE-associated loci were annotated to infantile epilepsy, microcephaly, severe learning difficulty, or other inborn diseases affecting the central nervous system (CNS) in OMIM (Supplementary Table 12).

Polymorphisms in genes linked to oculocutaneous albinism (OCA) were significantly associated with RE (Supplementary Table 13), although typically association was found for SNPs not strongly associated with other pigmentation traits[31]. Strong association with RE was found near the *OCA2* gene causing OCA type 2 (p=1.37x10^{-15}, rs79406658), OCA3 (*TYRP1*, p=1.18x10^{-11}, rs62538956), OCA5 (*SLC39A8*, p=4.03x10^{-17}, rs13107325), OCA6 (*C10orf11*, p=1.73x10^{-16}, rs12256171). In addition, significant association was found near genes linked to ocular albinism (OA) on chromosome X (*TBL1X* and *GPR143*[32], p=2.20x10^{-18}, rs34437079) and Hermansky-Pudlak Syndrome albinism (*BLOC1S1,* p=2.4610^{-22}, for rs80340147; note that this gene forms a conjoint read-through transcript the *BLOC1S1-RDH5* with *RDH5*). Other associated markers were located within genes involved in systemic pigmentation also previously associated with RE[7], such as *RALY* (p=3.14x10^{-18}, rs2284388), *TSPAN10* (p=2.22x^{-50}, rs9747347), as well as melanoma (*MCHR2,* p=2.37x10^{-15} for rs4839756).

Functional properties of the associated markers

Among the significantly associated markers, 367 unique markers were frameshift or missense variants (Supplementary Table 14). Several are non-synonymous, such as the R141L mutation (rs1048661) within *LOXL1,* a gene that causes pseudoexfoliation syndrome and glaucoma[33] and A69S (rs10490924) in *ARMS2*, associated with increased susceptibility to age-related macular degeneration[34]. Other associated variants with predicted deleterious consequences were located in several genes, such as RGR (p=6.89x10-68, rs1042454), a gene previously associated with RE[7,10] and also retinitis pigmentosa[35], and within the *FBN1* gene, near clusters of mutations that cause Marfan Syndrome and anterior segment dysgenesis[36].

Because the functional link between other associated variants and development of RE phenotypes is less obvious, we next performed gene-set enrichment analyses to identify properties that are significantly shared by genes identified by the meta-analysis. An enrichment analysis of Gene Ontology processes (Supplementary Table 15) found enrichment for genes participating in RNA Polymerase II transcription regulation ($p=1x10^{-06}$) and nucleic acid binding transcription factor activity ($p=1.10x10^{-06}$), suggesting that many of the genetic associations we identified interfere with gene expression. "Eye development" ($p=6.10x10^{-06}$) and "Circadian regulation of gene expression" ($p=1.10x10^{-04}$) were also significantly enriched.

A transcription factor binding site (TFBS) enrichment analysis identified significant (FDR < 0.05) over-representation of sites targeted by *GATA4*, *EP300*, *RREB1,* for which association was observed in the meta-analyses (Supplementary Table 16). Binding sites of transcription factors involved in eye morphogenesis and development such as *MAF* (whose mutations cause autosomal cataract), *FOXC1* and *PITX2* (anterior segment dysgenesis) or *CRX* (cone-rod dystrophy) were also enriched. *CRX* and *PAX4*, binding sites were also significantly enriched; these transcription factors are two of the regulators of circadian rhythm and melatonin synthesis[37] alongside *OTX2*, for which SNP significant association was observed in our RE meta-analysis ($p=6.15x10^{-11}$ for rs928109). All of these enriched gene-sets are observed for the first time in a GWAS analysis, although the presence of some of the mechanisms that relate them to RE and myopia were hypothesized before[38].

Many of the variants associated with RE in our analyses were located within or near genes that are expressed in numerous body tissues (Supplementary Figure 4), and in particular from the nervous system, consistent with our evidence of extraocular, central nervous system involvement in RE. Within the eye, these genes were particularly strongly expressed in eye tissues such as cornea, ciliary body, trabecular meshwork[39] and retina[40] (Supplementary Figure 5, Supplementary Table 17). A stratified LD score regression applied to specifically expressed genes (LDSC-SEG)[41] revealed the results of the GWAS are most strongly correlated with genes expressed in the retina and basal ganglia in the central nervous system but these correlations are not significant after multiple testing correction (Supplementary Figure 6 and Supplementary Table 18). It is possible that the strength of these correlations was constrained by the fact that in most cases, available expression levels were measured in adult samples, while refractive error and myopia are primarily developed in younger ages.

A Summary data-based Mendelian Randomization (SMR) analysis[42] integrating GWAS with eQTL data from peripheral blood[43] and brain tissues[44] found concomitant association with RE and eQTL transcriptional regulation effects for 159 and 97 genes respectively (Supplementary Tables 19 and 20). A similar analysis integrating GWAS summary data with methylation data from brain tissues found association with both RE and changes in methylation for 134 genes (Supplementary Table 21).

Pleiotropy and genetic effects shared between RE and other conditions

Examining the GWAS Catalog[45], some of the genetic variants reported here were previously associated with RE, and with other traits, in particular intraocular pressure, intelligence and education; the latter two are known myopia risk factors (Supplementary Table 22). We used LD score regression to assess the correlation of genetic effects between RE and other phenotypes from GWAS summary statistics (Supplementary Table 23). RE genetic risk was significantly correlated with intelligence, both in childhood[46] (r_g=-0.27, p=4.76x10^{-09}) and adulthood (fluid intelligence score r_g=-0.25, p=1.56x10^{-39}), educational attainment (defined as the number of years spent in formal education, r_g=-0.24, p=3.36x10^{-54}), self-reported cataract (r_g=-0.31, p=4.70x10^{-10}) and intraocular pressure (IOP, r_g=-0.14, p=1.04x10^{-12}).

Higher educational attainment appears to cause myopia as demonstrated by Mendelian randomization (MR) studies[47]. A gene by environment interaction GWAS for spherical equivalent and educational attainment (using age at completion of formal full-time education as a proxy) was conducted in 66,242 UK Biobank participants. Despite the relatively well-powered sample, only one locus yielded evidence of statistically significant interaction (rs536015141 within *TRPM1*, p=2.35x10^{-09}, Supplementary Table 24), suggesting that the true relationship between RE and education is compounded by several factors and may not be linear in nature, as suggested recently[48]. TRPM1 is localized in rod ON bipolar cell dendrites, and rare mutations cause congenital stationary night blindness[49], often associated with high myopia.

To further explore the nature of the relationship between RE and IOP, we built MR models using genetic effects previously reported for IOP[50]. On average, every 1 mmHg increase in IOP predicts a 0.05-0.09 diopters decrease in spherical equivalent (Supplementary Table 25, Supplementary Figure 7). We also built a MR model to assess the relationship between intelligence and spherical equivalent, but statistical evidence in this case points towards genetic pleiotropy rather than causation (Supplementary Table 26). This suggests that both myopia and intelligence are often influenced by the same factors, but without direct causal path linking one to the other. We found no significant genetic correlations between RE and the glaucoma endophenotype vertical cup to disc ratio (r_g=-0.01, p=0.45), or hair pigmentation (r_g=-0.03, p=0.35). Therefore, RE and pigmentation may have different allelic profiles with limited sharing of genetic risk.

Conditional analysis and risk prediction

We subsequently carried out a conditional analysis[51] on the meta-analysis summary results and found a total of 904 independent SNPs significantly associated with RE. 890 of these markers were available in the EPIC-Norfolk Study, an independent cohort that did not participate in the RE meta-analysis (Supplementary Figure 8). These markers alone explained 12.1% of the overall spherical equivalent phenotypic variance in a regression model or 18.4% (SE=0.04) of the spherical equivalent heritability. Newly associated markers found in our meta-analysis, but not in the previous large GWAS[7], explain 4.6% (SE=0.01) of the spherical equivalent phenotypic variance in EPIC-Norfolk Study, which is an improvement of one third compared to heritability explained by previously associated markers[7].

Predictive models, based on the above-mentioned 890 SNPs, along with age and sex, were predictive of myopia (versus all non-myopia controls) with areas under the receiving operating characteristic curve (AUC) of 0.67, 0.74 and 0.75 (Figure 2), depending on the severity cutoff for myopia ($\leq$ -0.75D, $\leq$ -3.00D and $\leq$ -5.00D respectively). The performance of the predictions appears not to improve for myopia definitions of -3.00D or worse, suggesting that the information extracted from our meta-analysis is more representative of the genetic risk for common myopia seen in the general population, than for more severe forms of myopia, which may have a distinctive genetic architecture.

Analysis of the distribution of effects and number of associated variants needed to explain all RE heritability

Using information from over half a million population-based participants SNPs identified in these analyses still only explain 18.4% of the spherical equivalent heritability. We next assessed how many common SNPs are likely to explain the entire heritable component of RE, and what sample sizes are likely to be needed in the future to identify them, using the likelihood-based approach described elswhere[52]. We estimate that approximately 13,808 (SE=969) polymorphic variants are likely to be behind the full RE heritability. Similar to other quantitative phenotypic traits that are previously published[52], our analyses estimate that 10.3% (SE=1.0%) of the phenotypic variance is likely explained by a batch of approximately 543 (SE=81) common genetic variants of relatively large effect size and a further 20.8% (SE=0.9%) of the entire phenotypic variance explained by the remainder. With increased sample sizes, we project that the proportion of variance explained will continue to improve fast but will start plateauing for sample sizes above one million, after which further increases in sample size will likely yield ever diminishing additional phenotypic variance (Supplementary Figure 9).

Discussion

Our results provide evidence for at least two major sets of mechanisms in the pathogenesis of RE. The first affect intraocular pressure, eye structure, ocular development and physiology, and the second are CNS-related, including circadian rhythm control. Contributors to RE include all anatomical factors that alter refractive power relative to eye size, light transmittance, photoconductance and higher cerebral functions.

The findings implicate almost every single anatomical components of the eye, which along with the central nervous system participate in the development of RE. The healthy cornea contributes to 70% of the optical refractive power of the eyes[53] and genes involved in corneal structure, topography and function may directly contribute to RE through direct changes in the corneal refraction. Our results show that several genes involved in lens development also contribute to RE in the general population. It is unclear if their contribution is mediated through alterations in biomechanical properties that affect eyes' ability to accommodate, changes to the lens refractive index, or alterations in light transmission properties that impair the ability to focus images on the retina.

Many retinal genes are implicated in the development of refractive error, reflecting the role of light in mediating eye growth and the importance of the retina's role in light transduction and processing[7]. Associations with RE at genes coding for gated ion channels and glutamate receptors point to the photoreceptor-bipolar cell interface as a potentially key factor in RE. Rare mutations in several of our associated genes cause night blindness, implicating the rod system in the pathophysiology of RE, but many also affect cone pathways. The *TRPM1* gene, important for rod ON bipolar cell polarity[54], is also implicated in the gene-education interaction analysis. Associations observed for the *VSX1* and VSX2, its negative regulator, genes implicate the cone bipolar cells[55].

The association with genes involved in pigmentation, including most of the OCA-causing genes, raises questions about the relationship between melanin, pigmentation and eye growth and development. These associations are unlikely to be influenced by any cryptic population structure in our samples, which our analyses were designed to control. None of the major pigmentation-associated SNPs[31] was directly associated with RE and there was no significant correlation of genetic effects between RE and pigmentation.

The mechanisms linking pigmentation with RE are unclear. Foveal hypoplasia[56] and optic disc[57] dysplasias are common in all forms of albinism[58]. Although melanin synthesis is disrupted in albinism, both melanin and dopamine are synthesized through shared metabolic pathways. Disc and chiasmal lesions in albinism are often attributed to dopamine[59], but we found limited evidence supporting an association with RE for genetic variants involved in dopamine signaling. The scarcity of association with RE for genes involved in dopamine-only pathways contrasts with the abundance of association for genes involved in pigmentation and melanin synthesis. This may suggest that melanin metabolism is connected to RE through other mechanisms that are independent from the metabolic pathways it shares with dopamine production. Melanin reaches the highest concentrations in the retinal pigment epithelium at the outmost layer of the retina, and anteriorly, in the iris and variations in pigmentation may affect the intensity of the light reaching the retina. Light exposure is a major protective factor for development of myopia[60,61] It is possible that pigmentation plays a role in light signal transmission and transduction.

Animal model experiments suggest that in addition to local ocular mechanisms, emmetropization (the process by which the eye develops to minimize refractive error) is strongly influenced by the CNS[62]. The strong correlation of genetic risks between RE and intelligence and association found for genes linked to severe learning disability support an involvement of the CNS in emmetropization and RE pathogenesis.

Results from gene-set enrichment analysis demonstrate an interesting evolution with increasing sample sizes. While smaller previous studies were sufficiently powered to discover enrichment of low, cell-level properties, such as cation channel activity and participation in the synaptic space structures[25], significantly more powered recent studies have found additional evidence for enrichment and involvement of more integrated physiological functions, such as light signal processing in retinal cells and others[7]. Beyond the identification of a much larger number of genes and explaining significantly higher proportions of heritability, our results, based in a considerably more statistically powered sample, uphold the previous findings and support the involvement of the same molecular and physiological mechanisms that were previously described.

In line with expectations from a higher power of association to discover genes and gene sets individually responsible for even smaller proportions of the refractive error variance[63], we find evidence for even higher regulatory mechanisms, that act more holistically over the eye development or integrate eye growth and homeostasis with other processes of extraocular nature. For example, we found evidence that binding sites of transcription factors involved in the control of circadian rhythm are significantly enriched among genes associated with refractive error. Circadian rhythm is important in emmetropization and its disruption leads to myopia in animal knock-out models[38], potentially through dopamine-mediated mechanisms, or changes in IOP and diurnal variations.

Most of the loci identified through our meta-analysis are not subject to particularly strong and systematic evolutionary pressures (Supplementary Figure 10). The variability in minor allele frequencies observed across loci associated with RE may therefore be the result of genetic drift. However, given the variety of the different visual components whose disruptions can result in RE, this variability may also be the result of overall balancing forces which encourage high allelic diversity of genes involved in RE, providing additional buffering capacity to absorb environmental pressures[48] or genetic disruptions on any of the individual components of the visual system.

Our results cast light on potential mechanisms that contribute to RE in the general population and have identified the genetic factors that explain a considerable proportion of the heritability and phenotypic variability of RE. This allows us to improve significantly our ability to make predictions of myopia risk and generate novel hypotheses on how multiple aspects of visual processing affect emmetropization, which may pave the way to personalized risk management and treatment of RE in the population in the future.

Online Methods provided separately

Figure 1. All GWAS-associated regions from the main meta-analysis. Each band is a true scale of genomic regions associated with refractive error listed in Supplementary Table 1 (+250kbp on each side to make smaller regions more visible). The different color codes represent the significance (p-value) for the genetic variant within that region that displays the strongest evidence for association.

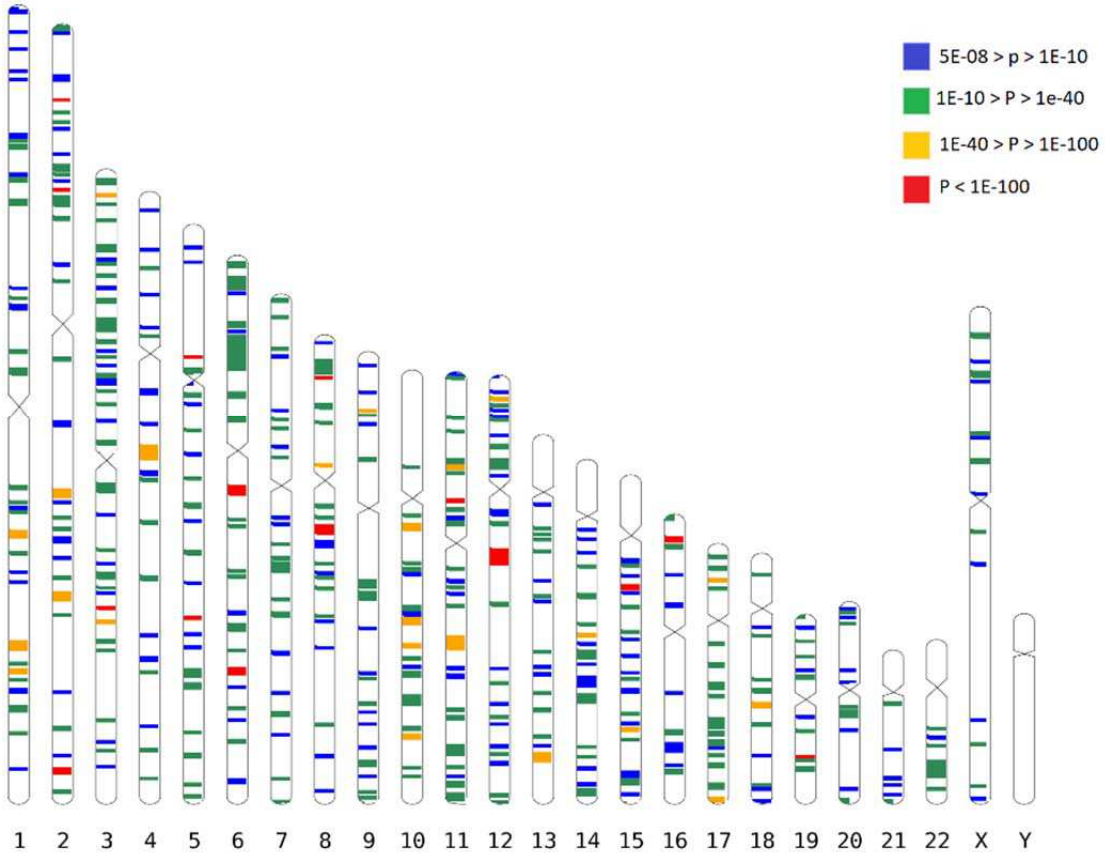
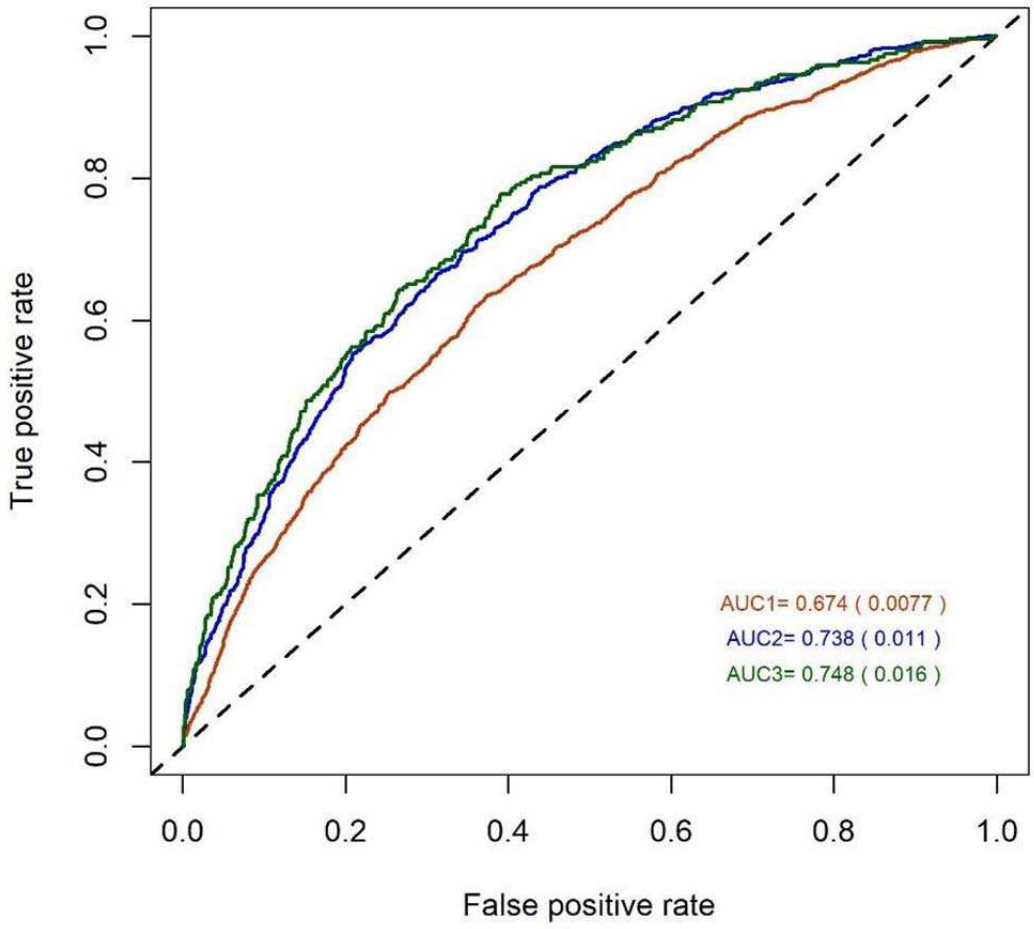


Figure 2. Receiver Operating Characteristic (ROC) curves for myopia predictions, using information from 890 SNP markers identified in the meta-analysis. The three different colors represent three different curves for each of the different definition of myopia: red – all myopia (< -0.75D), blue – moderate myopia (< -3.00 D) and green - severe myopia (defined as < -5.00 D).



References:

1. Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1211-1259 (2017).
2. WHO. The Global Burden of Disease. 2004 Update ISBN-13: 9789241563710. *ISBN-10* **651629118**(2008).
3. Williams, K.M. *et al.* Increasing Prevalence of Myopia in Europe and the Impact of Education. *Ophthalmology* **122**, 1489-97 (2015).
4. Sanfilippo, P.G., Hewitt, A.W., Hammond, C.J. & Mackey, D.A. The heritability of ocular traits. *Surv Ophthalmol* **55**, 561-83 (2010).
5. 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).
6. Verhoeven, V.J. *et al.* Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat Genet* **45**, 314-8 (2013).
7. Tedja, M.S. *et al.* Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat Genet* **50**, 834-848 (2018).
8. Cumberland, P.M. *et al.* Frequency and Distribution of Refractive Error in Adult Life: Methodology and Findings of the UK Biobank Study. *PLoS One* **10**, e0139780 (2015).
9. Kvale, M.N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1051-60 (2015).
10. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
11. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
12. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* **32**, 227-34 (2008).
13. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).
14. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).
15. Oliveira, J. *et al.* LAMA2 gene analysis in a cohort of 26 congenital muscular dystrophy patients. *Clin Genet* **74**, 502-12 (2008).
16. Colognato, H. *et al.* Identification of dystroglycan as a second laminin receptor in oligodendrocytes, with a role in myelination. *Development* **134**, 1723-36 (2007).
17. Burkin, D.J. & Kaufman, S.J. The alpha7beta1 integrin in muscle development and disease. *Cell Tissue Res* **296**, 183-90 (1999).
18. Ervasti, J.M. & Campbell, K.P. Dystrophin-associated glycoproteins: their possible roles in the pathogenesis of Duchenne muscular dystrophy. *Mol Cell Biol Hum Dis Ser* **3**, 139-66 (1993).
19. Mayer, U. *et al.* Absence of integrin alpha 7 causes a novel form of muscular dystrophy. *Nat Genet* **17**, 318-23 (1997).
20. Jean, D., Ewan, K. & Gruss, P. Molecular regulators involved in vertebrate eye development. *Mech Dev* **76**, 3-18 (1998).

21. Hammond, C.J., Andrew, T., Mak, Y.T. & Spector, T.D. A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: a genomewide scan of dizygotic twins. *Am J Hum Genet* **75**, 294-304 (2004).
22. Ali, M. *et al.* Null mutations in LTBP2 cause primary congenital glaucoma. *Am J Hum Genet* **84**, 664-71 (2009).
23. Clark, A.M. *et al.* Negative regulation of Vsx1 by its paralog Chx10/Vsx2 is conserved in the vertebrate retina. *Brain Res* **1192**, 99-113 (2008).
24. Heon, E. *et al.* VSX1: a gene for posterior polymorphous dystrophy and keratoconus. *Hum Mol Genet* **11**, 1029-36 (2002).
25. Hysi, P.G. *et al.* Common mechanisms underlying refractive error identified in functional analysis of gene lists from genome-wide association study results in 2 European British cohorts. *JAMA Ophthalmol* **132**, 50-6 (2014).
26. Ma, M. *et al.* Wnt signaling in form deprivation myopia of the mice retina. *PLoS One* **9**, e91086 (2014).
27. Miyake, M. *et al.* Identification of myopia-associated WNT7B polymorphisms provides insights into the mechanism underlying the development of myopia. *Nat Commun* **6**, 6689 (2015).
28. Cuellar-Partida, G. *et al.* WNT10A exonic variant increases the risk of keratoconus by decreasing corneal thickness. *Hum Mol Genet* **24**, 5060-8 (2015).
29. Stone, R.A. *et al.* Image defocus and altered retinal gene expression in chick: clues to the pathogenesis of ametropia. *Invest Ophthalmol Vis Sci* **52**, 5765-77 (2011).
30. Zhou, H., Yoshioka, T. & Nathans, J. Retina-derived POU-domain factor-1: a complex POU-domain gene implicated in the development of retinal ganglion and amacrine cells. *J Neurosci* **16**, 2261-74 (1996).
31. Hysi, P.G. *et al.* Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat Genet* **50**, 652-656 (2018).
32. Fabian-Jessing, B.K. *et al.* Ocular albinism with infertility and late-onset sensorineural hearing loss. *Am J Med Genet A* **176**, 1587-1593 (2018).
33. Thorleifsson, G. *et al.* Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* **317**, 1397-400 (2007).
34. Rivera, A. *et al.* Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum Mol Genet* **14**, 3227-36 (2005).
35. Morimura, H., Saindelle-Ribeaudeau, F., Berson, E.L. & Dryja, T.P. Mutations in RGR, encoding a light-sensitive opsin homologue, in patients with retinitis pigmentosa. *Nat Genet* **23**, 393-4 (1999).
36. Robinson, P.N. *et al.* Mutations of FBN1 and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. *Hum Mutat* **20**, 153-61 (2002).
37. Rohde, K., Moller, M. & Rath, M.F. Homeobox genes and melatonin synthesis: regulatory roles of the cone-rod homeobox transcription factor in the rodent pineal gland. *Biomed Res Int* **2014**, 946075 (2014).
38. Chakraborty, R. *et al.* Circadian rhythms, refractive development, and myopia. *Ophthalmic Physiol Opt* **38**, 217-245 (2018).
39. Carnes, M.U., Allingham, R.R., Ashley-Koch, A. & Hauser, M.A. Transcriptome analysis of adult and fetal trabecular meshwork, cornea, and ciliary body tissues by RNA sequencing. *Exp Eye Res* **167**, 91-99 (2018).
40. Ratnapriya, R. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat Genet* **51**, 606-610 (2019).

41. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
42. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
43. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
44. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* **9**, 2282 (2018).
45. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).
46. Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol Psychiatry* **19**, 253-8 (2014).
47. Mountjoy, E. *et al.* Education and myopia: assessing the direction of causality by mendelian randomisation. *BMJ* **361**, k2022 (2018).
48. Pozarickij, A. *et al.* Quantile regression analysis reveals widespread evidence for gene-environment or gene-gene interactions in myopia development. *Commun Biol* **2**, 167 (2019).
49. Audo, I. *et al.* TRPM1 is mutated in patients with autosomal-recessive complete congenital stationary night blindness. *Am J Hum Genet* **85**, 720-9 (2009).
50. Khawaja, A.P. *et al.* Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. *Nat Genet* **50**, 778-782 (2018).
51. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
52. Zhang, Y., Qi, G., Park, J.H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet* **50**, 1318-1326 (2018).
53. Zadnik, K. *et al.* Normal eye growth in emmetropic schoolchildren. *Optom Vis Sci* **81**, 819-28 (2004).
54. Li, Z. *et al.* Recessive mutations of the gene TRPM1 abrogate ON bipolar cell function and cause complete congenital stationary night blindness in humans. *Am J Hum Genet* **85**, 711-9 (2009).
55. Chow, R.L. *et al.* Vsx1, a rapidly evolving paired-like homeobox gene expressed in cone bipolar cells. *Mech Dev* **109**, 315-22 (2001).
56. Struck, M.C. Albinism: Update on ocular features. *Current Ophthalmology Reports* **3**, 232-237 (2015).
57. Mohammad, S. *et al.* Characterization of Abnormal Optic Nerve Head Morphology in Albinism Using Optical Coherence Tomography. *Invest Ophthalmol Vis Sci* **56**, 4611-8 (2015).
58. Yahalom, C. *et al.* Refractive profile in oculocutaneous albinism and its correlation with final visual outcome. *Br J Ophthalmol* **96**, 537-9 (2012).
59. Lopez, V.M., Decatur, C.L., Stamer, W.D., Lynch, R.M. & McKay, B.S. L-DOPA is an endogenous ligand for OA1. *PLoS Biol* **6**, e236 (2008).
60. Karouta, C. & Ashby, R.S. Correlation between light levels and the development of deprivation myopia. *Invest Ophthalmol Vis Sci* **56**, 299-309 (2014).
61. Wu, P.C., Tsai, C.L., Wu, H.L., Yang, Y.H. & Kuo, H.K. Outdoor activity during class recess reduces myopia onset and progression in school children. *Ophthalmology* **120**, 1080-5 (2013).
62. Troilo, D., Gottlieb, M.D. & Wallman, J. Visual deprivation causes myopia in chicks with optic nerve section. *Curr Eye Res* **6**, 993-9 (1987).
63. de Leeuw, C.A., Neale, B.M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat Rev Genet* **17**, 353-64 (2016).

515