Revisiting the Trans-Ancestry Genetic Correlation of Refractive Error

Rosie Clark[1], Xi He[1], Thu Nga Nguyen[1], Thanh Huyen Bui[1], Hannah Noor[1], Cathy Williams[2], Louise Terry[1], and Jeremy A. Guggenheim[1]; on behalf of the UK Biobank Eye and Vision Consortium

[1]School of Optometry & Vision Sciences, Cardiff University, Cardiff, United Kingdom
[2]Centre for Academic Child Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Correspondence: Jeremy A. Guggenheim, School of Optometry and Vision Sciences, Cardiff University, Cardiff CF24 4HQ, UK; guggenheimj1@cardiff.ac.uk.

Received: May 8, 2025
Accepted: July 31, 2025
Published: August 26, 2025

Citation: Clark R, He X, Nguyen TN, et al. Revisiting the trans-ancestry genetic correlation of refractive error. *Invest Ophthalmol Vis Sci.* 2025;66(11):60. https://doi.org/10.1167/iovs.66.11.60

Purpose. The prevalence of myopia varies significantly across the globe. This may be a consequence of differences in exposure to lifestyle risk factors or differences in genetic susceptibility across ancestry groups. "Trans-ancestry genetic correlation" quantifies the similarity in genetic predisposition to a trait or disease between different populations. We estimated the trans-ancestry genetic correlation of refractive error across Europeans, South Asians, East Asians, and Africans using recently developed approaches.

Methods. Two methods were applied: (1) trans-ancestry genetic correlation with unbalanced data resources (TAGC-UDR) and (2) trans-ancestry bivariate genomic-relatedness-based restricted maximum-likelihood (TAB-GREML). TAGC-UDR analyses were carried out for UK Biobank participants of European (n = 3500), East Asian (n = 972), South Asian (n = 4303), and African (n = 3877) ancestry. TAB-GREML analyses were carried out for participants of European (n = 10,000), South Asian (n = 4303), and African (n = 3877) ancestry.

Results. TAGC-UDR analyses suggested the trans-ancestry genetic correlation of refractive error was in the range 0.7–1.0 for the European versus African, European versus East Asian, and European versus South Asian ancestry pairs. The TAB-GREML estimates were consistent with the TAGC-UDR findings. Precision of the estimates was limited, reflecting the modest sample sizes of the non-European samples.

Conclusions. These results support and extend previous work suggesting that genetic susceptibility to refractive error is largely shared across Europeans, Africans, and South Asians. This suggests geographical differences in myopia prevalence are mostly driven by lifestyle factors or rare genetic variants not considered in the current work.

Keywords: myopia, refractive error, molecular genetics, ancestry, genomics

Refractive errors occur because of a mismatch between the combined focal power of the cornea and crystalline lens compared to the eye's axial length.[1] In myopia, the axial length is too long relative to the eye's optical power, which causes light to focus in front of the plane of the retinal photoreceptors. The converse happens in hyperopia, with light focusing behind the plane of the retinal photoreceptors. The blurry vision produced by myopia and hyperopia can be corrected by wearing glasses or contact lenses or by refractive surgery. However, myopia and hyperopia are associated with serious comorbidities.[2] For instance, myopia is a risk factor for glaucoma, retinal detachment, and myopic macular degeneration, which are increasingly common causes of visual impairment.[2–4]

Both genetics and environmental (lifestyle) factors contribute to the development of refractive errors.[2] In certain individuals, early-onset high myopia is caused by a mutation in a single gene. To date, loss of function mutations in approximately 20 different genes have been found to cause non-syndromic high myopia, with *ARR3* and *OPN1LW* being the most frequently reported disease genes.[5–8] More generally, genetic susceptibility to myopia and hyperopia is conferred by inheritance of a very large number of genetic variants; each variant confers a very small increased risk of either myopia or hyperopia, yet in aggregate the combined effect of all of the variants is substantial.[9] More than 450 distinct genetic variants associated with refractive error have been identified by genome-wide association studies.[10] Currently, the cumulative effect of these genetic variants can explain about 20% of the intersubject variation in refractive error.[11] Insufficient time spent outdoors during childhood and a high level of education are the most important lifestyle risk factors for myopia that have been discovered.[12,13] Mounting evidence suggests that these lifestyle risk factors trigger the development of myopia in genetically susceptible individuals.[2] However, because the prevalence of myopia in young adults exceeds 80% in some countries, it is likely that even a person

with limited genetic susceptibility to myopia will eventually succumb if their exposure to lifestyle risk factors is high enough.[2,14]

A notable feature of the worldwide prevalence of myopia is its geographical variation, exceeding 50% in countries such as China, Taiwan, Hong Kong, and Singapore, while being below 20% in Australia and parts of Africa.[15,16] The varied geographical distribution of myopia may result from differences in genetic susceptibility across ancestry groups or differences in the level of exposure to lifestyle risk factors across geographical regions, or a combination of the two. "Trans-ancestry genetic correlation" is a statistical measure used to quantify the similarity in genetic predisposition to a trait or disease across two populations of differing ancestry. By analogy with a conventional correlation coefficient, a trans-ancestry genetic correlation can be conceptualized as a graph of effect sizes plotted for two separate populations. Each data point corresponds to a specific genetic variant; its position on the graph is determined by the variant's effect size in ancestry A (x-axis) and its effect in ancestry B (y-axis). The correlation coefficient for all points on the graph corresponds to the genetic correlation for the two populations. A trans-ancestry genetic correlation equal to one implies that, across the genome, genetic variants that confer susceptibility to a disease are shared perfectly between the two ancestral groups. In other words, a trans-ancestry genetic correlation of one implies that genetic variants have the same effect size (i.e., the same degree of association with the disease) in the two ancestral populations. By contrast, a trans-ancestry genetic correlation of zero implies unique sources of genetic susceptibility in the two populations. In practice, calculating a measure of trans-ancestry genetic correlation is complicated by differences in allele frequency across ancestry groups (in the extreme case, an allele that confers susceptibility to a disease may be present in one ancestry group but absent from the other). One approach that can be useful in these circumstances is to compare the local patterns of linkage disequilibrium (LD) in the ancestry groups, which enables the effects of an index variant to be inferred using information from adjacent variants.[17]

A range of statistical methods have been developed to estimate trans-ancestry genetic correlations; each method makes different assumptions regarding the effect of allele frequency differences or in its requirement for summary data or individual-level data.[17–22] In past work, the CREAM consortium reported a trans-ancestry genetic correlation for refractive error in European ancestry versus East Asian ancestry individuals,[23] using an analysis method called *popcorn*.[19] Here, we apply two recently introduced methods for calculating the trans-ancestry genetic correlation of refractive error that rely on different assumptions to *popcorn*. For our primary analysis we apply the Trans-Ancestry Genetic Correlation with Unbalanced Data Resources (TAGC-UDR) method,[17] which is designed for use when the sample size of one of the two ancestral populations is limited. As a sensitivity analysis we apply the trans-ancestry bivariate genomic-relatedness-based restricted maximum-likelihood (TAB-GREML) method, which fits a random effects model similar to the widely-employed bivariate GREML approach,[24] but that takes into account differences in allele frequency between the two ancestry samples rather than assuming a common allele frequency.[22,25] We build on the existing CREAM findings by estimating the trans-ancestry genetic correlation of refractive error for a diverse set of ancestry groups: European, East Asian, South Asian, and African.

Methods

Overview of the Study Methods

We applied two methods of calculating trans-ancestry genetic correlations: TAGC-UDR and TAB-GREML. The TAGC-UDR method (Fig. 1) was used as our primary analysis because this approach has been reported to perform well when one ancestry sample is much larger than the other.[17] TAB-GREML was used as a sensitivity analysis; this approach requires relatively large sample sizes for both ancestry groups and can be biased if the sample sizes are very highly imbalanced.[17]

TAGC-UDR. Unrelated UK Biobank participants with information available for refractive error were classified into one of five ancestry groups: European (n = 111,168), East Asian (n = 972), South Asian (n = 4303), African (n = 3877), and Other. The European sample was split into a genome-wide association study (GWAS) sample (n = 107,668) and an "evaluation sample" (n = 3500). A GWAS for refractive error was performed in the European ancestry GWAS sample (n = 107,668), using a subset of HapMap3 genetic variants.[26] The regression coefficients for all variants included in the GWAS were used to derive weights for a polygenic score, which was then applied to participants in the four evaluation samples: European (n = 3500), East Asian (n = 972), South Asian (n = 4303), and African (n = 3877). Trans-ancestry "bias factors"[17] were computed using LD reference samples from each ancestry group combined with an estimate of the single nucleotide polymorphism (SNP) heritability (h^2_{SNP}) of refractive error in each ancestry group (because SNP heritability is challenging to calculate in small samples, calculations were performed using a range of plausible values: h^2_{SNP} = 0.2–0.5). The role of the bias factors is to up-scale the polygenic score in the evaluation sample to account for (i) imprecision in GWAS beta coefficient estimation due to the finite size of the GWAS analysis, (ii) imprecision in estimating the accuracy of the polygenic score because of the small size of the non-European sample, and (iii) differences in LD-structure between ancestry groups[17] (Fig. 1 and Supplementary Note S1). Comparison of the true refractive error versus the up-scaled genetically predicted refractive error for participants in the evaluation samples was used to estimate the trans-ancestry genetic correlation. Standard errors (SE) were estimated by bootstrapping.

TAB-GREML. TAB-GREML analysis is a modification of standard bivariate GREML analysis[24] that takes into account the differences in allele frequency in two ancestry samples when calculating trans-ancestry genetic correlations.[22,25] Unrelated UK Biobank participants with information available for refractive error were classified into one of five ancestry groups as described in the TAGC-UDR section above. Next, a set of 10,000 European ancestry participants was selected at random from the full European-ancestry sample. TAB-GREML analyses were carried out for (a) n = 10,000 European and n = 4303 South Asian ancestry participants, and (b) the n = 10,000 European and n = 3877 African ancestry participants. The sample size of East Asian ancestry participants (n = 972) was too small to fit a TAB-GREML model, which prevented a European–East Asian trans-ancestry

Investigative Ophthalmology & Visual Science

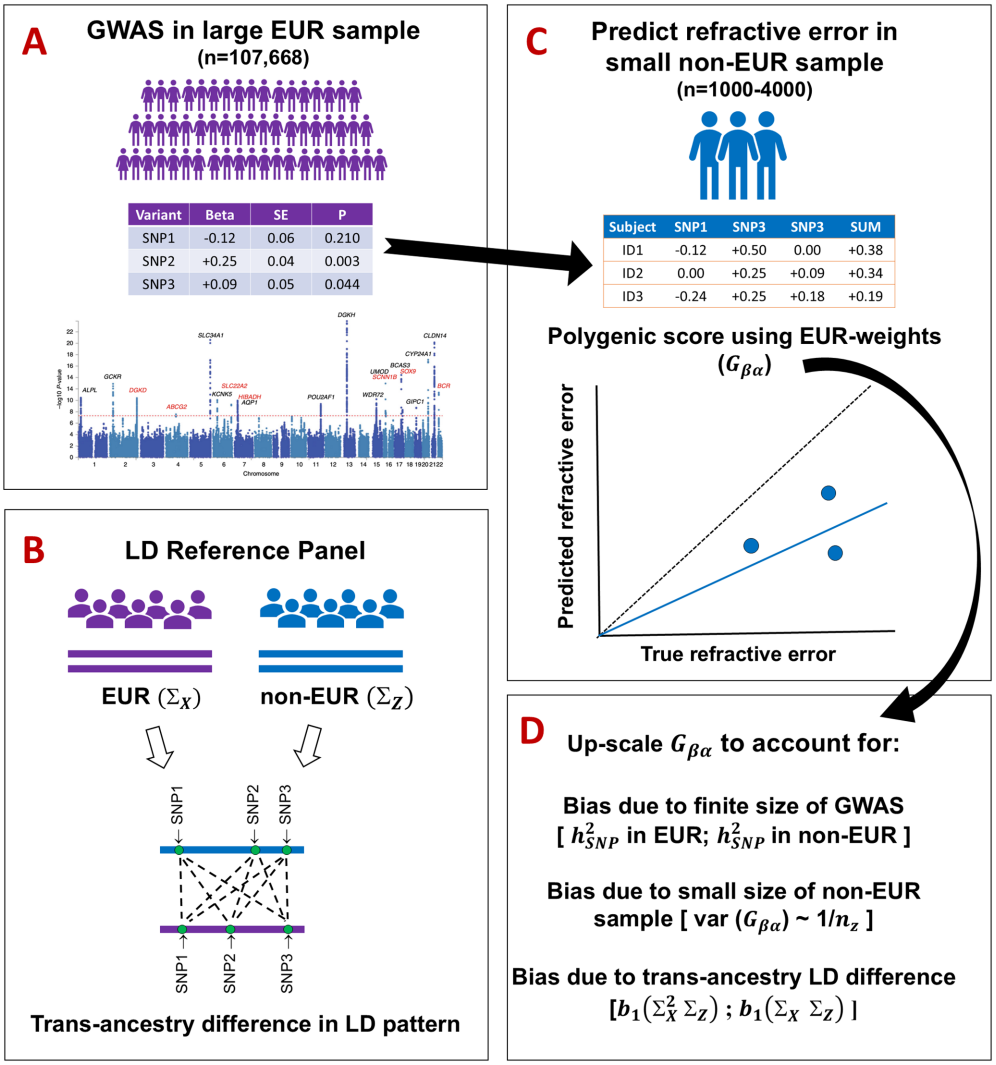

Figure 1. Schematic diagram of the trans-ancestry genetic correlation with unbalanced data resources (TAGC-UDR) method. **(A)** A GWAS for refractive error is carried out in a large sample of European (EUR) ancestry participants. **(B)** Differences in the pattern of LD across the genome are compared in LD blocks of EUR and non-EUR reference samples. For the p genetic variants in each LD block, the $p \times p$ LD matrices are termed Σ_X and Σ_Z for genetic variants x and z in the EUR and non-EUR reference samples, respectively. **(C)** The GWAS summary statistics from the EUR sample are used to derive weights ($\hat{\beta}$) for a polygenic score. The polygenic score is applied to the non-EUR evaluation sample (n = 1000-4000) to predict the refractive error of each participant. If the polygenic score had perfect accuracy, then the trans-ancestry genetic correlation ($G\beta\alpha$) could be calculated by comparing the true versus predicted refractive error in the non-EUR sample (where α is the vector of effect sizes for genetic variants z in the non-EUR population). **(D)** In practice, three major sources of bias must be accounted for to obtain an accurate estimate of $G\beta\alpha$. The first source of bias (#1) introduces noise when estimating genetic variant effect sizes ($\hat{\beta}$) in the EUR population, due to the finite size of the GWAS in the EUR sample. The second source of bias (#2) introduces noise when estimating genetic variant effect sizes ($\hat{\alpha}$) in the non-EUR population, because of the small size of the non-EUR evaluation sample. This bias is assumed to be proportional to the reciprocal of the sample size of the non-EUR evaluation sample, n_z. The third source of bias (#3) occurs when estimating $\hat{\alpha}$ using the polygenic score, due to LD differences between the EUR and non-EUR populations: $b_1(\Sigma^2_X\ \Sigma_Z)$ and $b_1(\Sigma_X\ \Sigma_Z)$, where b_1 is the limiting spectral distribution (LSD) of $\Sigma^{1/2}_X\ \Sigma^{1/2}_Z$. The *solid arrows* indicate the flow of information during the TAGC-UDR analysis.

genetic correlation being calculated using the TAB-GREML method.

Selection of Participants From the UK Biobank Cohort

Ethical approval for the UK Biobank study was obtained from the NHS Research Ethics Committee (Reference: 11/NW/0382). Approximately 500,000 participants aged 40–70 years-old were recruited from across the United Kingdom, during the period 2006 and 2010, via one of 22 assessment centers.[27] Approximately one quarter of the UK Biobank cohort had their refractive error measured (Tomey RC 5000 autorefractor; Tomey Corp., Nagoya, Japan). Blood samples were collected and genotyped at ~800,000 markers using either the UK BiLEVE Axiom array or the UK Biobank Axiom array. Imputation to a joint Haplotype Reference Consortium–UK 10,000 Genomes Project reference panel was performed by Bycroft et al.[28] Participants were included in the current analysis if they had information available for

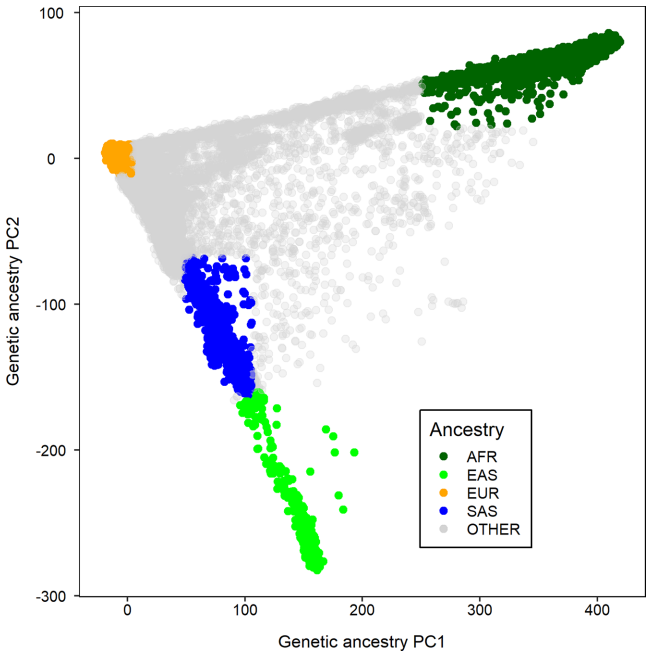

Figure 2. Genetic ancestry classification based on genetic principal components 1 and 2 (PC1 and PC2). In comparison to EUR participants, wider distributions of PC1 and PC2 were used to define AFR, EAS and SAS participants to maximize the available sample sizes. AFR, African; EAS, East Asian; EUR, European; SAS, South Asian.

their refractive error, were seen at an assessment center at which more than $n = 100$ participants had refractive error assessed, and had a genetic heterozygosity within 10 standard deviations of the mean for the full sample. Ancestry was classified using a loose definition based solely on genetic principal components 1 and 2 (PC1 and PC2) with the aim of maximizing the sample size of each ancestry group. Specifically, Europeans were defined as individuals with PC1 and PC2 within 10 standard deviations of the mean for participants who self-reported their ethnicity as "White British." East Asians were defined as individuals with PC1 and PC2 within five standard deviations of the mean for participants who self-reported their ethnicity as "Chinese." South Asians were defined as individuals with PC1 and PC2 within 1.25 standard deviations of the mean for participants who self-reported their ethnicity as "Asian." Africans were defined as individuals with PC1 and PC2 within 2 SD of the mean for participants who self-reported their ethnicity as "Black." Figure 2 illustrates the distribution of PC1 and PC2 in each ancestry group. Within each of the African, East Asian, and South Asian groups, a maximal set of unrelated participants was selected using a custom R script that applied the R package *igraph* function to the kinship data reported by Bycroft et al.[28] This resulted in samples of unrelated African ($n = 3877$), East Asian ($n = 972$), and South Asian ($n = 4303$) participants, which we refer to as "evaluation samples." For the European ancestry sample, a set of $n = 3500$ unrelated participants was first selected at random as a European "evaluation sample." The remaining $n = 107{,}668$ European ancestry participants were used as a GWAS sample. From within the African, East Asian, European, and South Asian evaluation samples, a set of approximately 1000 unrelated participants was selected at random to serve as an LD reference sample for that ancestry group.

Selection of Genetic Variants

The set of $n = 990{,}761$ HapMap3 variants reported by Zhao et al.[17] (https://github.com/FSSKM/TAGC_review/blob/main/TAGC_data/hm3_snp.id.list) was filtered to remove (i) variants with a Hardy Weinberg equilibrium test $P < 1e\text{-}04$ in any of the four LD reference ancestry samples, (ii) variants with a minor allele frequency (MAF) < 1% in any of the four LD reference ancestry samples, and (iii) variants with a genotyping rate < 90% in any of the four LD reference ancestry samples. This resulted in a final set of $n = 642{,}807$ HapMap3 variants.

Genome-Wide Association Study for Refractive Error

A GWAS for autorefractor-measured spherical equivalent refractive error (in diopters) averaged between the two eyes was performed using BOLT-LMM v2.4.1[29] for the $n = 107{,}668$ European ancestry GWAS sample. Age, age^2, sex, genotyping array, assessment center, and genetic ancestry PC1-10 were included as covariates. The $n = 642{,}807$ HapMap3 variants were analyzed.

Trans-Ancestry Genetic Correlation Analysis With TAGC-UDR

Using the terminology from Zhao et al.,[17] the $b_1(\Sigma_X^2\ \Sigma_Z)$ and $b_1(\Sigma_X^2\ \Sigma_Z)$ LD block matrix moment terms were calculated for each of the 253 LD blocks identified by Zhao et al.[17] in each ancestry pair (African-European, East Asian-European, South Asian-European, European-European) with the *TAGC_LD_block_moment* and *WAGC_LD_block_moment* functions from the R package *TAGC* (https://github.com/FSSKM/TAGC_review). These LD moments were then pooled across the genome for each ancestry pair and used to calculate trans-ancestry bias factors for each ancestry pair.[17]

A polygenic score for refractive error was calculated using PLINK v1.90b6[30] for participants in each of the four evaluation samples, using the SNP weights from the GWAS for autorefractor-measured spherical equivalent refractive error described above. The polygenic score and the autorefractor-measured refractive error were both standardized (transformed to have mean = 0, $SD = 1$) within each evaluation sample. Then, standardized autorefractor-measured refractive error was regressed on the standardized polygenic score, and a set of standardized covariates (age, age^2, sex, and PC1-PC10) using 500 bootstrap replicates. The regression coefficient and its bootstrap standard error were both multiplied by the ancestry-specific bias factor to yield an estimate of the trans-ancestry genetic correlation.[17]

Sensitivity Analysis: Trans-Ancestry Genetic Correlation Analysis With TAB-GREML

Sets of unrelated UK Biobank participants of African ($n = 3877$) and South Asian ($n = 4303$) ancestry were selected as described in the *Selection of participants from the UK Biobank cohort* section above. From the full set

of European-ancestry participants (n = 111,168), n = 10,000 unrelated participants were chosen at random. A genetic relationship matrix (GRM) was constructed for the combined African + European sample (n = 13,877) and for the combined South Asian + European sample (n = 14,303), using the *-rtmx2* command of the MTG2 software.[22] TAB-GREML analysis was performed using the *-mod 2* command of MTG2, with autorefractor-measured refractive error as the phenotype and age, age^2, sex, and PC1-PC10 as covariates.

Results

Demographic characteristics of the European-ancestry GWAS sample and the four multi-ancestry evaluation samples are presented in Table 1. All participants were adults aged 40–70 years. The African, East Asian and South Asian evaluation samples were all younger, on average, than the European evaluation sample (Mann-Whitney U-test; all $P < 0.001$). The median level of refractive error in the East Asian evaluation sample was negative, while it was positive in the other samples. Indeed, refractive error was significantly more myopic in the East Asian sample compared to the other three evaluation samples (Mann-Whitney U-test; all $P < 0.001$). In accordance with the more myopic refractive error in the East Asian sample, this ancestry group's self-reported age of first wearing glasses/contact lenses was younger than that of the other three ancestry samples (Mann-Whitney U-test; all $P < 0.001$). These ancestry-specific refractive error differences have been observed previously in UK Biobank participants.[31]

Figure 3 shows the trans-ancestry genetic correlation of refractive error across pairs of ancestry groups calculated using the TAGC-UDR method (further details can be found in Supplementary Fig. S1). The European versus European comparison was included as a calibration factor: an evaluation of participants with the same genetic ancestry would be expected to have a genetic correlation of $r_g = 1.0$. Estimation of the trans-ancestry genetic correlation requires the "SNP-heritability" of refractive error in each ancestry sample to be specified (SNP-heritability refers to the genetic contribution to a trait that can be explained by a particular set of SNPs).[32] In Figure 3, SNP-heritability was assumed to be equal in each pair of ancestry groups being evaluated; this assumption was relaxed for the analyses shown in Supplementary Figure S1. Because a precise SNP-heritability value is challenging to estimate in small samples[24] and can vary depending on the specific set of SNPs being considered, we calculated the trans-ancestry genetic correlations across a range of plausible heritability values. At each potential SNP-heritability value, the estimated trans ancestry genetic correlation was nominally higher in the European versus European ancestry pair, followed by the European versus South Asian ancestry pair, European versus East Asian ancestry pair and finally the European versus African ancestry pair (Fig. 3). However, the 95% confidence intervals of these pairwise estimates overlapped, suggesting that the trans-ancestry genetic correlation estimates were similar, given the level of precision achieved by the TAGC-UDR analysis method.

When the SNP-heritability of refractive error was assumed to be $h^2_{SNP} = 0.2$, the trans-ancestry genetic correlation esti-

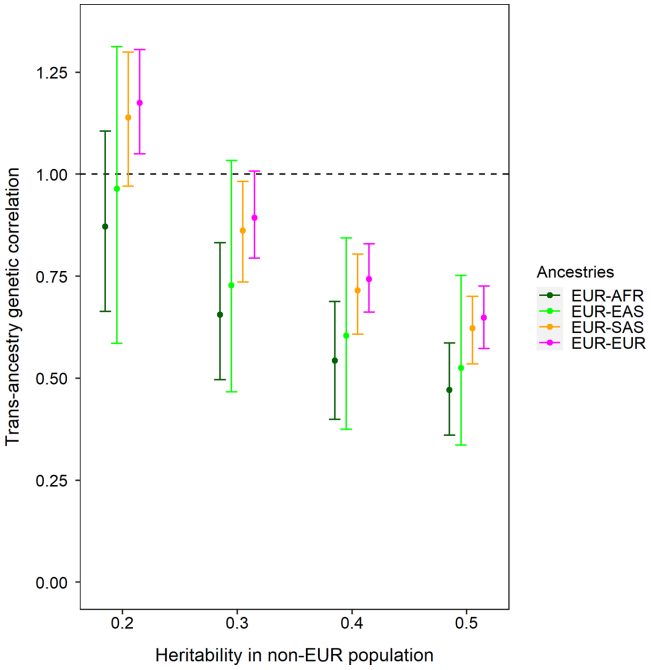

Figure 3. Trans-ancestry genetic correlation estimates. Error bars correspond to 95% confidence intervals. The EUR-EUR comparison was included as a calibration factor, with the expectation that this would provide a genetic correlation $r_g \approx 1.0$. AFR, African; EAS, East Asian; EUR, European; SAS, South Asian.

Table 1. Demographic Characteristics of the GWAS Sample and the Multi-Ancestry Evaluation Samples

Evaluation Sample	European GWAS Sample	African Evaluation Sample	East Asian Evaluation Sample	European Evaluation Sample	South Asian Evaluation Sample
Sample size	107,668	3877	972	3500	4303
Female percentage	53.4%	59.1%	68.0%	54.1%	48.2%
Age (years) median (Q1–Q3)	59.83 (52.17–64.50)	51.67 (46.08–58.58)	53.83 (47.17–60.42)	60.00 (52.24–64.83)	54.17 (46.71–61.33)
Age of first spectacles (years)* median (Q1–Q3)	37.00 (15.00–47.00)	37.50 (19.00–45.00)	25.00 (14.00–45.00)	38.00 (15.00–48.00)	40.00 (18.00–45.00)
Refractive error (D) median (Q1–Q3)	0.14 (−1.23 to 1.14)	0.04 (−0.82 to 0.70)	−0.44 (−2.98 to 0.45)	0.16 (−1.18 to 1.14)	0.02 (−1.05 to 0.76)

* Age of first spectacles had a bi-modal distribution with peaks at around 13 years-old and 50 years-old, reflecting correction for myopia and presbyopic, respectively.

Table 2. Trans-Ancestry Genetic Correlation Estimates for Refractive Error

Analysis Method	Ancestry Pair	Heritability ± SE in European Sample*	Heritability ± SE in Non-European Sample*	Trans-Ancestry Genetic Correlation ± SE†
TAGC-UDR	European vs. African	0.200	0.200	0.872 ± 0.109
	European vs. African	0.300	0.300	0.656 ± 0.084
	European vs. African	0.400	0.400	0.543 ± 0.071
	European vs. African	0.500	0.500	0.471 ± 0.059
	European vs. East Asian	0.200	0.200	0.964 ± 0.193
	European vs. East Asian	0.300	0.300	0.728 ± 0.143
	European vs. East Asian	0.400	0.400	0.604 ± 0.119
	European vs. East Asian	0.500	0.500	0.525 ± 0.103
	European vs. South Asian	0.200	0.200	1.139 ± 0.081†
	European vs. South Asian	0.300	0.300	0.862 ± 0.063
	European vs. South Asian	0.400	0.400	0.715 ± 0.049
	European vs. South Asian	0.500	0.500	0.622 ± 0.043
TAB-GREML	European vs. African	0.237 ± 0.030	0.302 ± 0.087	0.489 ± 0.181
	European vs. South Asian	0.241 ± 0.030	0.275 ± 0.062	1.072 ± 0.173†

* For TAGC-UDR, heritability was input as a model parameter; for TAB-GREML, heritability was estimated from the data.
† Note that TAGC-UDR and TAB-GREML do not constrain genetic correlation estimates to lie within the range −1.0 to 1.0, because this can introduce bias.

mates were close to $r_g \approx 0.9$–1.0 for all ancestry pairs: $r_g = 1.14$, 95% confidence interval (CI) = 0.97–1.30 for the European versus South Asian ancestry pair; $r_g = 0.96$, 95% CI = 0.59–1.31 for the European versus East Asian pair; $r_g = 0.87$, 95% CI = 0.66 to 1.11 for the European versus African ancestry pair. (Note that in order to provide unbiased estimates, the TAGC-UDR method permits trans-ancestry genetic correlations to fall outside the range 0–1, even though such values are not possible in reality). When the SNP-heritability of refractive error was assumed to be higher, the trans-ancestry genetic correlation estimates fell; at the highest SNP heritability level of $h^2_{SNP} = 0.5$, the trans-ancestry genetic correlation was estimated to be $r_g \approx 0.5$–0.6 for all ancestry pairs: $r_g = 0.62$, 95% CI = 0.54–0.70 for the European versus South Asian ancestry pair; $r_g = 0.53$, 95% CI = 0.34–0.75 for the European versus East Asian pair; $r_g = 0.47$, 95% CI = 0.36 to 0.59 for the European versus African ancestry pair. In Figure 3, a genetic correlation of $r_g \approx 1.0$ for the European versus European ancestry pair—matching that expected for the calibration factor comparison—was observed for SNP-heritability values $h^2_{SNP} = 0.2$–0.3. For this SNP-heritability level, the trans-ancestry genetic correlation of refractive error was in the range $r_g = 0.7$–1.0 for the European versus African, European versus East Asian and European versus South Asian ancestry pairs (Table 2).

As a sensitivity analysis, the trans-ancestry genetic correlation for the European versus African and European versus South Asian ancestry pairs was calculated using an alternative method, TAB-GREML. As shown in Table 2, the TAB-GREML estimates were $r_g = 0.49$ (SE = 0.18) for the European versus African ancestry pair and $r_g = 1.07$ (SE = 0.17) for the European versus South Asian ancestry pair. The high trans-ancestry genetic correlation for the European versus South Asian ancestry pair supported the result obtained using TAGC-UDR. The TAB-GREML trans-ancestry genetic correlation for the European versus African ancestry pair was lower than that estimated with TAGC-UDR (TAB-GREML: $r_g = 0.49$; TAGC-UDR: $r_g = 0.66$–0.87; Table 2). However, these estimates all had large standard errors, suggesting the differences could either be real or have arisen because of lack of precision.

Discussion

Our analyses suggested the trans-ancestry genetic correlation of refractive error most likely lay within the range $r_g = 0.7$–1.0 across a wide range of ancestry groups—Europeans, Africans, East Asians and South Asians – implying that geographic variation in the prevalence of myopia is mainly driven by differences in lifestyle rather than genetics. The results are in line with the high genetic correlation of refractive error between European and East Asian individuals reported by the CREAM consortium ($r_g = 0.8$–0.9), which was calculated using different sets of participants to those studied here and using an alternative method.[23] Our work extends the earlier CREAM findings by estimating the trans-ancestry genetic correlation across a wider range of ancestries; to our knowledge, the current study is the first to reveal that genetic susceptibility to refractive error is largely shared across Europeans, Africans, and South Asians.

When performing a within-ancestry genetic correlation analysis—for example, the correlation of SNP effect sizes for the traits refractive error and axial length in European samples[33]—the allele frequency of the SNPs with phenotypic effects will generally be well matched for any two samples of European ancestry. By contrast, in a trans-ancestry genetic correlation analysis, many SNPs may be common in one ancestry group but rare in the other. A SNP that has the same effect size but differs in allele frequency across two ancestral populations will have a greater impact on the phenotype in the ancestry group with the higher allele frequency.[18] Brown et al.[19] coined the terms "trans-ethnic genetic-impact correlation" (ρ_{gi}) and "trans-ethnic genetic-effect correlation" (ρ_{ge}) to distinguish between genetic correlation estimates that do or do not account for ancestry-specific allele frequency differences, respectively. The TAGC-UDR and TAG-GREML methods we used here produce estimates that are more akin to ρ_{ge} than ρ_{gi}.

A key advantage of the TAGC-UDR method for estimating trans-ancestry genetic correlations is that a large sample of participants is only required for one ancestry group (usually, this will be a group of European-ancestry participants, due to the greater availability of Biobank-scale

resources for Europeans). Thus the method can be applied when the available sample size of the second ancestry groups is limited to a few hundred or a few thousand participants.[17] A disadvantage of the TAGC-UDR method is that it produces imprecise estimates, which is a consequence of the small sample size of the second ancestry group. Furthermore, although the TAGC-UDR method accounts for differences in LD patterns across ancestry groups genome-wide, this information is summarized as a set of three LD block moment values (one of which corresponds to the number of SNPs included in the analysis). This approach, while shown to be unbiased in simulations,[17] may not capture subtle trans-ancestry differences in LD structure. Following Momin et al.,[22] we restricted the HapMap3 SNPs used in our analysis to those with a MAF of at least 1% in the ancestry groups studied to avoid the situation where an allele was essentially monomorphic in one ancestry sample. This meant that our genetic correlation was based on a selected sample of genetic variants, whereas ideally a genetic correlation estimate would be calculated using the set of all causal variants in the genome for the trait-of-interest. Rare genetic variants (those with MAF <1%) tend to have arisen in populations more recently than common variants (MAF ≥1%). Thus a trans-ancestry genetic correlation calculated for common plus rare SNPs would likely be lower than a trans-ancestry genetic correlation calculated for just common SNPs. The SNP-heritability of refractive error in Europeans calculated using HapMap3 SNPs is 0.25–0.45.[32,34] Because of the limited sample size of the evaluation samples, we calculated genetic correlations across a range of plausible heritability values, rather than estimating the heritability directly from the data. As expected theoretically,[17] trans-ancestry genetic correlation estimates were lower when the heritability of refractive error was assumed to be higher (Fig. 3). This phenomenon occurs because the TAGC-UDR method scales-up SNP effect sizes based on the h^2_{SNP} (Fig. 1): for instance, if the true h^2_{SNP} = 0.3 but it is incorrectly assumed to be h^2_{SNP} = 0.5, then insufficient up-scaling will occur, which will result in a trans-ancestry genetic correlation that is lower than the true value. Conversely, if the true h^2_{SNP} = 0.3 but it is incorrectly assumed to be h^2_{SNP} = 0.2, then excessive up-scaling will occur, resulting in an inflated estimate of the true trans-ancestry genetic correlation. A disadvantage of our approach of estimating genetic correlations across a range of plausible heritability values was that this left open the question of which heritability value is most appropriate. We addressed this by selecting the heritability value (h^2_{SNP} = 0.2–0.3) that provided the theoretically expected genetic correlation for the European vs. European pair comparison.

The European sample in our analyses had a median age that was six to eight years older than the other ancestry samples (Table 1). In the absence of a cohort effect, this difference in age would be unlikely to have had a major impact on our results, because (i) we included age as a covariate in our analyses, (ii) by adulthood, both the genetic and environmental contribution to refractive error will have become fully manifest, and (iii) refractive error is relatively stable across the 40- to 70-year age range.[35] In reality, cohort effects on the refractive error of UK Biobank participants seem likely to be present, especially in relation to a trend towards greater educational attainment for those born in more recent decades.[36–38] Nevertheless, although heritability will vary across birth cohorts if environmental influences vary over time, we would expect a six- to eight-year age gap to have minimal influence. Thus GWAS regression coefficients would be largely stable across the age range of UK Biobank participants. The East Asian sample in our analysis had a median refractive error that was more negative (myopic) than that of the other ancestry samples. However, this would have been unlikely to affect our genetic correlation estimates since heritability and genetic correlation are based on within-sample variations in trait values (i.e., changing the mean value of a trait does not alter the heritability or genetic correlation estimates for the trait).

In summary, genetic correlations for refractive error across European-African, European-East Asian, and European-South Asian ancestry pairs were estimated to lie in the range r_g = 0.7–1.0. The current result for European vs. East Asian ancestry individuals matched that obtained by the CREAM consortium using an alternative method. The relatively high trans-ancestry genetic correlations suggest that the wide variation in the prevalence of myopia across geographic regions is mainly due to differences in exposure to lifestyle risk factors rather than genetics. However, a caveat to this conclusion is that the current study focused on common variants and therefore does not rule out the possibility that rare variants make an important contribution to geographic variation in the prevalence of myopia. Furthermore, small sample sizes for the non-European ancestry samples limited the precision of the current genetic correlation estimates, which indicates a need for Biobank-scale collection of genotyped individuals with known refractive error across diverse ancestries. A recent consensus statement from the International Myopia Summit[39] stressed that public health strategies aimed at reducing visual impairment caused by myopia need to have a global perspective. Our findings lend support to the idea that public health efforts targeting lifestyle risk factors for myopia will be effective across a wide range of geographical regions.

Acknowledgments

This research has been conducted using the UK Biobank Resource (application no. 83325). UK Biobank was established by the Wellcome Trust; the UK Medical Research Council; the Department for Health (London, UK); Scottish Government (Edinburgh, UK); and the Northwest Regional Development Agency (Warrington, UK).

Supported by Project Grant (24WG201) from the Welsh Government and Fight for Sight; UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number EP/Y032292/1]. Funded by the European Union (Project 101119501 — MyoTreat — HORIZON-MSCA-2022-DN-01). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or UKRI. Neither the European Union nor the granting authority can be held responsible for them. It also received funding from the Welsh Assembly Government (Cardiff, UK); the British Heart Foundation; and Diabetes UK. Collection of eye and vision data was supported by The Department for Health through an award made by the NIHR to the Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust, and UCL Institute of Ophthalmology, London, United Kingdom (grant no. BRC2_009). Additional support was provided by The Special Trustees of Moorfields Eye Hospital, London, United Kingdom (grant no. ST 12 09). Data analysis was carried out using the HAWK computing clus-

ter, maintained by Supercomputing Wales and Cardiff University ARCCA.

Disclosure: **R. Clark**, None; **X. He**, None; **T.N. Nguyen**, None; **T.H. Bui**, None; **H. Noor**, None; **C. Williams**, None; **L. Terry**, None; **J.A. Guggenheim**, None

References

1. Baird PN, Saw S-M, Lanca C, et al. Myopia. *Nat Rev Dis Primers.* 2020;6:99.
2. National Academies of Sciences Engineering and Medicine. *Myopia: Causes, Prevention, and Treatment of an Increasingly Common Disease.* Washington, DC: The National Academies Press; 2024.
3. Sankaridurg P, Tahhan N, Kandel H, et al. IMI impact of myopia. *Invest Ophthalmol Vis Sci.* 2021;62:2.
4. Ohno-Matsui K, Wu P-C, Yamashiro K, et al. IMI pathologic myopia. *Invest Ophthalmol Vis Sci.* 2021;62:5.
5. Gardner JC, Liew G, Quan YH, et al. Three different cone opsin gene array mutational mechanisms; genotype-phenotype correlation and functional investigation of cone opsin variants. *Hum Mutat.* 2014;35:1354–1362.
6. Xiao X, Li S, Jia X, Guo X, Zhang Q. X-linked heterozygous mutations in ARR3 cause female-limited early onset high myopia. *Mol Vis.* 2016;22:1257–1266.
7. Wang Y, Xiao X, Li X, et al. Genetic and clinical landscape of ARR3-associated MYP26: the most common cause of Mendelian early-onset high myopia with a unique inheritance. *Br J Ophthalmol.* 2023;107:1545–1553.
8. Wang Y, Sun W, Xiao X, et al. Unique haplotypes in OPN1LW as a common cause of high myopia with or without protanopia: A potential window into myopic mechanism. *Invest Ophthalmol Vis Sci.* 2023;64:29.
9. Tideman JWL, Pärssinen O, Haarman AEG, et al. Evaluation of shared genetic susceptibility to high and low myopia and hyperopia. *JAMA Ophthalmol.* 2021;139:601–609.
10. Hysi PG, Choquet H, Khawaja AP, et al. Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. *Nat Genet.* 2020;52:401–407.
11. Clark R, Lee SS-Y, Du R, et al. A new polygenic score for refractive error improves detection of children at risk of high myopia but not the prediction of those at risk of myopic macular degeneration. *EBioMed.* 2023;91:104551.
12. Yang Y, Liao H, Zhao L, et al. Green space morphology and school myopia in China. *JAMA Ophthalmol.* 2024;142:115–122.
13. Morgan IG, Wu P-C, Ostrin LA, et al. IMI risk factors for myopia. *Invest Ophthalmol Vis Sci.* 2021;62:3.
14. Wojciechowski R. Nature and nurture: The complex genetics of myopia and refractive error. *Clin Genet.* 2011;79:301–320.
15. Attebo K, Ivers RQ, Mitchell P. Refractive errors in an older population: The Blue Mountains Eye Study. *Ophthalmol.* 1999;106:1066–1072.
16. Holden BA, Fricke TR, Wilson DA, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmol.* 2016;123:1036–1042.
17. Zhao B, Yang X, Zhu H. Estimating trans-ancestry genetic correlation with unbalanced data resources. *J Am Stat Assoc.* 2024;119:839–850.
18. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. *Bioinformatics.* 2012;28:2540–2542.
19. Brown BC, Ye CJ, Price AL, Zaitlen N. Transethnic genetic-correlation estimates from summary statistics. *Am J Hum Genet.* 2016;99:76–88.
20. Cai M, Xiao J, Zhang S, et al. A unified framework for cross-population trait prediction by leveraging the genetic correlation of polygenic traits. *Am J Hum Genet.* 2021;108:632–655.
21. Mester R, Hou K, Ding Y, et al. Impact of cross-ancestry genetic architecture on GWASs in admixed populations. *Am J Hum Genet.* 2023;110:927–939.
22. Momin MM, Shin J, Lee S, et al. A method for an unbiased estimate of cross-ancestry genetic correlation using individual-level data. *Nat Commun.* 2023;14:722.
23. Tedja MS, Wojciechowski R, Hysi PG, et al. Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat Genet.* 2018;50:834–848.
24. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88:76–82.
25. Guo J, Bakshi A, Wang Y, et al. Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Sci Rep.* 2021;11:5240.
26. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467:52–58.
27. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12:e1001779.
28. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203–209.
29. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet.* 2015;47:284–290.
30. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4:7.
31. Cumberland PM, Bao Y, Hysi PG, et al. Frequency and distribution of refractive error in adult life: Methodology and findings of the UK Biobank Study. *PLoS ONE.* 2015;10:e0139780.
32. Guggenheim JA, St Pourcain B, McMahon G, et al. Assumption-free estimation of the genetic contribution to refractive error across childhood. *Mol Vis.* 2015;21:621–632.
33. Jiang C, Melles RB, Yin J, et al. A multiethnic genome-wide analysis of 19,420 individuals identifies novel loci associated with axial length and shared genetic influences with refractive error and myopia. *Front Genet.* 2023;14:1113058.
34. Shah RL, Guggenheim JA, U.K. Biobank Eye and Vision Consortium. Genome-wide association studies for corneal and refractive astigmatism in UK Biobank demonstrate a shared role for myopia susceptibility loci. *Human Genetics.* 2018;137:881–896.
35. Weale RA. Epidemiology of refractive errors and presbyopia. *Surv Ophthalmol.* 2003;48:515–543.
36. Williams KM, Bertelsen G, Cumberland P, et al. Increasing Prevalence of Myopia in Europe and the Impact of Education. *Ophthalmol.* 2015;122:1489–1497.
37. Cumberland PM, Bountziouka V, Hammond CJ, Hysi PG, Rahi JS. Temporal trends in frequency, type and severity of myopia and associations with key environmental risk factors in the UK: Findings from the UK Biobank Study. *PLoS ONE.* 2022;17:e0260993.
38. Plotnikov D, Williams C, Atan D, et al. Effect of education on myopia: Evidence from the United Kingdom ROSLA 1972 Reform. *Invest Ophthalmol Vis Sci.* 2020;61:7.
39. Furukawa T, Morrow EM, Cepko CL. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell.* 1997;91:531–541.

Investigative Ophthalmology & Visual Science