MHJ
LUMORA JOURNALS

Multidisciplinary Healthcare Journal
Vol. 1, No. 6 (2025) · e-ISSN 3152-8775

Original Article · DOI: not registered

Transferability of European-Ancestry Blood Lipid Loci to a Middle Eastern Arab Population: A Power-Calibrated Cross-Ancestry Replication, Summary-Statistic Polygenic Score and Fine-Mapping Study Using 2021 Qatar Biobank Whole-Genome and Global Lipids Genetics Consortium Summary Statistics

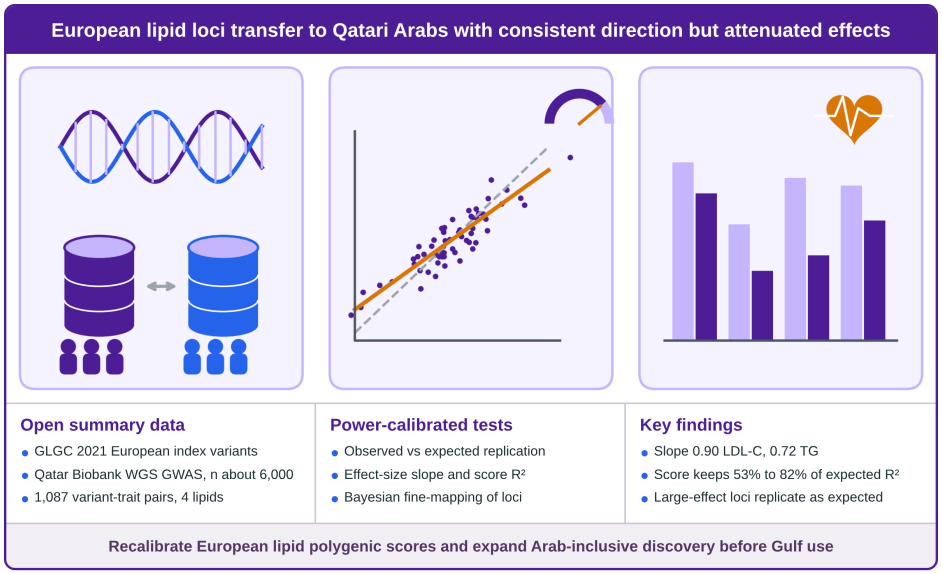
Hadeel Rshash Almutairi[1,*], Haila Alshelowi[2], Hyder Mirghani[3], Mohammed Alhassan[4], Najlaa Al-Mana[5], Zahwah Dahmi Alshammari[6]

Received	Acceptance date	Publication date
20 June 2025	25 June 2025	30 June 2025

[1] Qassim University at Unizah College, Saudi Arabia
[2] Qassim Armed Forces Hospital, Qassim, Saudi Arabia
[3] University of Tabuk, Tabuk, Saudi Arabia
[4] Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia
[5] King Saud bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia
[6] University of Ha'il, Saudi Arabia

*** Correspondence:** fatemalumora@gmail.com

ABSTRACT



GRAPHICAL ABSTRACT

Background: Most lipid-associated genetic variants were discovered in European-ancestry cohorts, and Arab populations remain under-represented. Earlier Gulf studies reported low replication of these variants but did not compare the observed replication with the replication expected from the statistical power of the smaller Arab studies. We quantified how well European-ancestry lipid loci transfer to Qatari Arabs.

Methods: This cross-ancestry genetic association study used public summary statistics only. European-ancestry index variants for low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG) and total cholesterol (TC) were taken from the Global Lipids Genetics Consortium 2021 (n = 1,320,016). They were matched to the 2021 Qatar Biobank whole-genome sequencing genome-wide association study (n = 5,972 to 6,017). For each trait we estimated directional concordance, nominal and Bonferroni replication against power-based expectations, an inverse-variance-weighted effect-size slope, between-ancestry heterogeneity and the variance explained by a European-weighted score. We fine-mapped overlapping genome-wide significant Qatari loci with Wakefield approximate Bayes factors under a single-causal-variant assumption. Reporting followed STREGA.

Results: Of 1,600 catalogued index variants, 1,087 variant-trait pairs were analysable. Directional concordance ranged from 69.5% (LDL-C) to 75.1% (HDL-C), below the 77.3% to 80.2% expected under full transfer. Nominal replication was lower than expected for LDL-C (29 vs 42.9; ratio 0.68), TG (23 vs 44.7; ratio 0.51) and TC (36 vs 47.6; ratio 0.76), whereas Bonferroni-significant replication matched expectation (5 to 7 variants per trait). Qatari effects were attenuated relative to European effects for TC (slope 0.88, 95% CI 0.78–0.97), HDL-C (0.78, 0.68–0.88) and TG (0.72, 0.64–0.81); for LDL-C the slope was 0.90 (0.80–1.00), with an interval that reached 1. The European-weighted score explained 6.46% of LDL-C variance (relative to expectation 0.82, 95% CI 0.66–0.99), 5.63% of TC (0.77), 4.13% of TG (0.53) and 3.57% of HDL-C (0.60). Among 16 single-signal Qatari loci, the median 95% credible set contained 6 variants; APOE rs7412 and GCKR rs1260326 resolved to single variants. Five low-frequency multi-signal loci on chromosome 11q23 were excluded from fine-mapping; their pattern is consistent with, but does not establish, an extended Qatari haplotype. These score estimates assume independent variants and are not individual-level validation; TC attenuation was not retained at 5 Mb distance pruning.

Conclusions: European-ancestry lipid loci showed directionally concordant associations in Qatari Arabs, with lower effect-size slopes for HDL-C and TG. Summary-derived score variance estimates are conditional on independence and harmonised trait scales; they are not validated individual-level predictive performance. Findings support Arab-inclusive discovery and local validation before clinical use.

Keywords: genome-wide association study; lipids; polygenic score; transferability; fine-mapping; Qatar; Arabs

Cite as: Almutairi HR, Alshelowi H, Mirghani H, Alhassan M, Al-Mana N, et al. Transferability of European-Ancestry Blood Lipid Loci to a Middle Eastern Arab Population: A Power-Calibrated Cross-Ancestry Replication, Summary-Statistic Polygenic Score and Fine-Mapping Study Using 2021 Qatar Biobank Whole-Genome and Global Lipids Genetics Consortium Summary Statistics. Multidisciplinary Healthcare Journal. 2025;1(6):MHJ-2026-000374. © The authors. Published open access under the Creative Commons Attribution 4.0 International licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

1. Introduction

Atherosclerotic cardiovascular disease is the leading cause of death worldwide, and elevated low-density lipoprotein cholesterol (LDL-C) is a causal risk factor [1,2]. In Saudi Arabia and the wider Gulf, dyslipidaemia is common from adolescence onwards, and national guidelines now place it at the centre of cardiovascular prevention [3,4]. Genetic information could support this agenda, through polygenic scores that identify people at high lifetime risk and through causal variants that point to drug targets. Most of that genetic knowledge, however, comes from studies of European ancestry [5,6]. Polygenic scores built from such studies lose accuracy in other ancestries, which risks widening health disparities [7,35].

The Global Lipids Genetics Consortium (GLGC) meta-analysed 1.65 million individuals in 2021. European-ancestry participants dominated the meta-analysis, and the consortium showed that adding diverse ancestries improved fine-mapping and score portability [8]. Loss of portability across ancestries has several sources: differences in allele frequency, differences in linkage disequilibrium (LD) between tag and causal variants, and genuine differences in effect size [9–12]. These sources have practical consequences. If weak replication in a new population reflects only low statistical power, European loci can be used as they are. If it reflects attenuated effects, scores need to be recalibrated locally.

Arab populations of the Arabian Peninsula have distinct population histories, high rates of consanguinity and long runs of homozygosity [13–15]. The Qatar Genome Programme released whole-genome sequencing association results for 45 clinical traits, including four blood lipids [16,17]. Earlier assessments of transferability to the Arabian Peninsula concluded that only a small proportion of reported metabolic variants replicated [18]. A recent Qatari LDL-C study replicated 168 GLGC loci and reported a moderate correlation of effect sizes [19], and polygenic scores for cardiometabolic disease have been evaluated in Saudi Arabs [20]. Saudi genome-wide studies have identified loci for myocardial infarction and hypertriglyceridaemia [21,22]. None of these studies asked how much replication should be expected given the power of an Arab cohort of about 6,000 people. None combined, across all four standard lipid traits, a power-calibrated replication analysis, an effect-size calibration slope, a summary-level estimate of polygenic score transfer and statistical fine-mapping on fully open data.

We therefore assessed the transferability of GLGC 2021 European-ancestry index variants for LDL-C, high-density lipoprotein cholesterol (HDL-C), triglycerides (TG) and total cholesterol (TC) to the Qatari population, using only public summary statistics. Our objectives were to (i) compare observed directional concordance and replication with power-based expectations; (ii) estimate the attenuation of effect sizes and its consequence for the variance explained by a European-weighted polygenic score; (iii) test for between-ancestry heterogeneity at individual loci; and (iv) fine-map genome-wide significant Qatari loci that overlap European signals.

2. Methods

2.1 Study design and reporting

This was a cross-ancestry genetic association study based on secondary analysis of published genome-wide association study (GWAS) summary statistics. It was reported in accordance with the STrengthening the REporting of Genetic Association Studies (STREGA) extension of STROBE [23]. The analysis plan (traits, matching rules, metrics, thresholds and fine-mapping prior) was fixed before outcome data were examined and is summarised in the protocol available with the study materials.

2.2 Data sources

Qatari summary statistics came from the Qatar Genome Programme GWAS of 45 clinical traits [16], conducted in participants of the population-based Qatar Biobank [17]. The discovery analysis included 6,218 whole-genome-sequenced adults and used linear mixed models adjusted for age, sex, principal components and relatedness. Traits were rank-based inverse-normal transformed, so effects are expressed in standard deviation (SD) units per allele [16]. We downloaded the harmonised GRCh38 files for LDL-C (GCST90013333; n = 5,972), HDL-C (GCST90013332; n = 6,013), TG (GCST90013334; n = 6,017) and TC (GCST90013331; n = 6,017) from the NHGRI-EBI GWAS Catalog [24] on 6 November 2026 and verified each file against the repository checksum.

European-ancestry index variants came from the GLGC 2021 European-ancestry analyses (n = 1,320,016) [8], retrieved through the GWAS Catalog interface for accessions GCST90239658 (LDL-C), GCST90239652 (HDL-C), GCST90239664 (TG) and GCST90239676 (TC). GLGC residuals were inverse-normalised, so the European effects are also in SD units [8]. For each index variant we recorded the GRCh38 position, effect allele, effect-allele frequency, per-allele effect, 95% confidence interval (CI) and P value. The standard error (SE) was derived as the CI width divided by 3.92.

2.3 Variant matching and quality control

Qatari variants with minor allele frequency (MAF) below 0.01 were excluded, consistent with the source analysis. Index variants were matched to Qatari variants by chromosome, position and allele. Qatari effects and frequencies were aligned to the European effect allele. Strand-ambiguous (A/T or C/G) variants with Qatari MAF above 0.42 were excluded. When a position was duplicated, the record with the smallest European P value was kept. The Qatari files report effects and P values without SEs, so the SE was recovered as $|\beta|/z$, where z is the standard normal quantile of $1 - P/2$. When P exceeded 0.92 and this ratio became unstable, the SE was approximated by $1/\sqrt{(2f(1 - f)N)}$, where f is the effect-allele frequency and N the sample size. Genomic control λ was estimated from a systematic 1-in-100 sample of all analysable Qatari variants.

2.4 Outcomes and statistical analysis

All analyses were performed separately for each lipid trait.

Directional concordance and replication. A variant was concordant when its Qatari effect had the same sign as its European effect. It replicated nominally when, in addition, Qatari $P < 0.05$, and it showed Bonferroni replication when Qatari $P < 0.05/m$, where m is the number of analysed variants for that trait. Expected values assumed full transfer, that is, a true Qatari effect equal to the European effect. Under this assumption, the probability of concordance for variant i is $\Phi(|\beta_{E,i}|/SE_{Q,i})$ and the power for replication is $\Phi(|\beta_{E,i}|/SE_{Q,i} - z_{crit})$. Here Φ is the standard normal distribution function, $\beta_{E,i}$ the European effect, $SE_{Q,i}$ the Qatari SE and z_{crit} the two-sided critical value. Observed counts were compared with the Poisson-binomial distribution of these probabilities, using the lower-tail probability. Proportions are reported with Clopper–Pearson 95% CIs.

Effect-size calibration. Qatari effects were regressed on European effects through the origin, weighted by $1/SE_Q^2$ [25]. The slope equals 1 under full transfer and falls below 1 when effects are attenuated. Its SE was inflated by the square root of the residual dispersion when the dispersion exceeded 1. The Pearson correlation of effects is reported with a Fisher-transformed 95% CI.

Between-ancestry heterogeneity. For each variant we computed Cochran's Q statistic comparing the European and Qatari effects (1 degree of freedom), its I^2 statistic [26] and a fixed-effect inverse-variance meta-analytic estimate.

Summary-statistic polygenic score transfer. A score weighted by the European effects was evaluated from summary statistics after pruning index variants to at least 500 kb apart. Assuming independent variants and a standardised trait, the covariance between score and trait is $\Sigma w_i \beta_{Q,i} H_i$ and the score variance is $\Sigma w_i^2 H_i$, where $w_i = \beta_{E,i}$ and $H_i = 2f_{Q,i}(1 - f_{Q,i})$. The variance explained (R^2) equals the squared covariance divided by the

score variance. We subtracted the sampling variance of the covariance, $\Sigma(w_i H_i)^2 SE_{Q,i}^2$, from the squared covariance to remove noise-induced upward bias, and derived 95% CIs from the normal CI of the covariance. Expected R^2 under full transfer equals the score variance. We computed it with Qatari frequencies and with European frequencies. The ratio of the two isolates the effect of allele-frequency differences, and the ratio of observed to expected R^2 (relative R^2) isolates the effect of attenuated effect sizes.

Qatari genome-wide significant loci and fine-mapping. Qatari variants with $P < 5 \times 10^{-8}$ were clumped by distance (500 kb) into lead variants, and each lead was checked for an index variant of any lipid trait within 500 kb. Each Qatari lead lying within 250 kb of a same-trait European index variant was fine-mapped across ±250 kb under a single-causal-variant model. We used Wakefield approximate Bayes factors with a prior effect SD of 0.15 [27], derived posterior inclusion probabilities (PIP) and formed 95% credible sets [28]. Prior SDs of 0.10 and 0.20 were used as a sensitivity analysis. Loci with another Qatari genome-wide significant lead within 1 Mb were flagged as multi-signal regions and excluded from the fine-mapping summary.

Multiplicity and correlated traits. LDL-C and TC are strongly correlated, and many index variants are shared across traits. We therefore report each trait separately and treat pooled counts across traits as descriptive only. Bonferroni thresholds were defined within each trait, and no formal test pools the four traits.

Sensitivity analyses. The calibration slope was re-estimated (a) restricted to index variants with Qatari MAF ≥ 0.05, (b) restricted to index variants with European $P < 5 \times 10^{-12}$, which limits winner's-curse effects, and (c) with every Qatari SE replaced by the frequency-based approximation.

Additional distance-pruning sensitivity was performed post hoc during editorial review; see Results.

Software. Analyses used Python 3.13.18 with NumPy 2.5.3, pandas 2.3.3, SciPy 1.18.1 and statsmodels 0.15.0. Figures were drawn with matplotlib 3.11.2. A single reproducible script reads only the frozen extracts.

3. Results

3.1 Data and variant matching

The Qatari GWAS provided 7,731,946 analysable variants per trait. Genomic control λ ranged from 0.990 to 1.056 (Table 1). Of 1,600 catalogued European index variants (403 for LDL-C, 380 for HDL-C, 388 for TG and 429 for TC), 4 lacked a parseable effect estimate, confidence interval or position, 494 were not present in the Qatari data, mostly because they were rare or absent there, and 15 strand-ambiguous variants were excluded. This left 269, 261, 274 and 283 variants, respectively (1,087 variant-trait pairs). Of the 509 parseable index variants not analysed, 250 had a European MAF below 0.05 and 132 a European MAF below 0.01; their median European MAF was 0.054, compared with a median Qatari MAF of about 0.25 among analysed variants. The analysed set therefore consists mainly of common variants. The median Qatari MAF of the analysed variants ranged from 0.244 to 0.274, and effect-allele frequencies were strongly correlated between ancestries (r = 0.92 to 0.93; Figure 3A). The Qatari SE had to be approximated for 4.5% to 7.4% of variants. Where both SE estimates were available, the two agreed closely (r ≥ 0.999). Because most European effects were small (median absolute European effect 0.015 to 0.016 SD per allele), the median power for nominal replication of an individual variant in the Qatari sample was only 0.097 to 0.115.

Table 1. Data sources, variant matching and quality indicators for the four lipid traits.

Characteristic	LDL-C	HDL-C	TG	TC
Qatari GWAS sample size	5,972	6,013	6,017	6,017
Qatari variants analysed (MAF ≥ 0.01)	7,731,946	7,731,946	7,731,946	7,731,946
Genomic-control λ (Qatari GWAS)	0.990	1.056	1.016	0.998
European index variants catalogued	403	380	388	429
Not available in Qatari data	132	111	110	141

Characteristic	LDL-C	HDL-C	TG	TC
Excluded, ambiguous palindromic	1	7	4	3
Index variants analysed	269	261	274	283
Median Qatari MAF of analysed variants	0.246	0.274	0.266	0.244
Allele-frequency correlation, r (95% CI)	0.92 (0.90–0.93)	0.92 (0.90–0.94)	0.93 (0.91–0.94)	0.92 (0.90–0.93)
Median power for nominal replication	0.099	0.115	0.101	0.097

GWAS, genome-wide association study; MAF, minor allele frequency; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglycerides; TC, total cholesterol; CI, confidence interval. European index variants are the Global Lipids Genetics Consortium 2021 European-ancestry associations (n = 1,320,016). Power is the probability of $P < 0.05$ in the same direction if the Qatari effect equalled the European effect. Four catalogued index variants (LDL-C 1, HDL-C 1, TC 2) could not be parsed (missing effect, confidence interval or position), so the catalogued count exceeds the sum of the next three rows by these variants.

3.2 Directional concordance and power-calibrated replication

Across traits, 790 of 1,087 variant-trait pairs (72.7%) had the same direction of effect in both ancestries. Concordance ranged from 69.5% for LDL-C to 75.1% for HDL-C (Table 2, Figure 2A), far above chance but below the 77.3% to 80.2% expected under full transfer. The shortfall was statistically detectable for LDL-C ($P = 0.001$), HDL-C ($P = 0.023$) and TG ($P = 0.024$), but not for TC ($P = 0.060$).

Nominal replication occurred for 129 pairs in total, against 184.3 expected. By trait, 29 of 269 LDL-C variants (10.8%) replicated against 42.9 expected (observed-to-expected ratio 0.68; $P = 0.004$), 41 of 261 HDL-C variants (15.7%) against 49.1 (ratio 0.83; $P = 0.082$), 23 of 274 TG variants (8.4%) against 44.7 (ratio 0.51; $P < 0.001$) and 36 of 283 TC variants (12.7%) against 47.6 (ratio 0.76; $P = 0.017$) (Figure 2B). By contrast, Bonferroni-significant replication matched its expectation for every trait: 5 vs 5.9 for LDL-C, 5 vs 6.4 for HDL-C, 7 vs 7.8 for TG and 6 vs 7.3 for TC (all $P \geq 0.32$; Figure 2C). Variants with large effects therefore transferred as expected, whereas the deficit came from the many variants with small effects.

Table 2. Transferability of European-ancestry lipid index variants to the Qatari population.

Metric	LDL-C	HDL-C	TG	TC
Same direction, n (%)	187 (69.5)	196 (75.1)	199 (72.6)	208 (73.5)
Expected same direction, %	77.3	80.2	77.7	77.5
P, observed vs expected direction	0.001	0.023	0.024	0.060
Nominal replication, observed/expected	29/42.9	41/49.1	23/44.7	36/47.6
Observed-to-expected ratio, nominal	0.68	0.83	0.51	0.76
P, observed vs expected nominal	0.004	0.082	< 0.001	0.017
Bonferroni replication, observed/expected	5/5.9	5/6.4	7/7.8	6/7.3
Effect-size slope (95% CI)	0.90 (0.80–1.00)	0.78 (0.68–0.88)	0.72 (0.64–0.81)	0.88 (0.78–0.97)
Effect-size correlation, r (95% CI)	0.77 (0.72–0.82)	0.53 (0.44–0.62)	0.69 (0.63–0.75)	0.74 (0.68–0.79)
Heterogeneity P < 0.05, n (%)	21 (7.8)	19 (7.3)	18 (6.6)	22 (7.8)
Heterogeneity after Bonferroni, n	0	1	1	0
Score R^2 expected, European frequencies (%)	9.81	5.86	7.32	8.28
Score R^2 expected, Qatari frequencies (%)	7.91	5.90	7.85	7.35
Score R^2 observed (%) (95% CI)	6.46 (5.25–7.82)	3.57 (2.75–4.52)	4.13 (3.23–5.17)	5.63 (4.51–6.92)

Metric	LDL-C	HDL-C	TG	TC
Relative R^2 (observed/expected) (95% CI)	0.82 (0.66–0.99)	0.60 (0.47–0.77)	0.53 (0.41–0.66)	0.77 (0.61–0.94)

Expected values assume that each Qatari per-allele effect equals the European effect, given the Qatari standard error. The slope is the inverse-variance-weighted regression of Qatari on European effects through the origin (1 indicates full transfer). Heterogeneity is Cochran's Q with 1 degree of freedom per variant. Score R^2 is the variance in each lipid trait explained by a European-weighted score of the index variants, estimated from summary statistics assuming independent variants. Abbreviations as in Table 1.

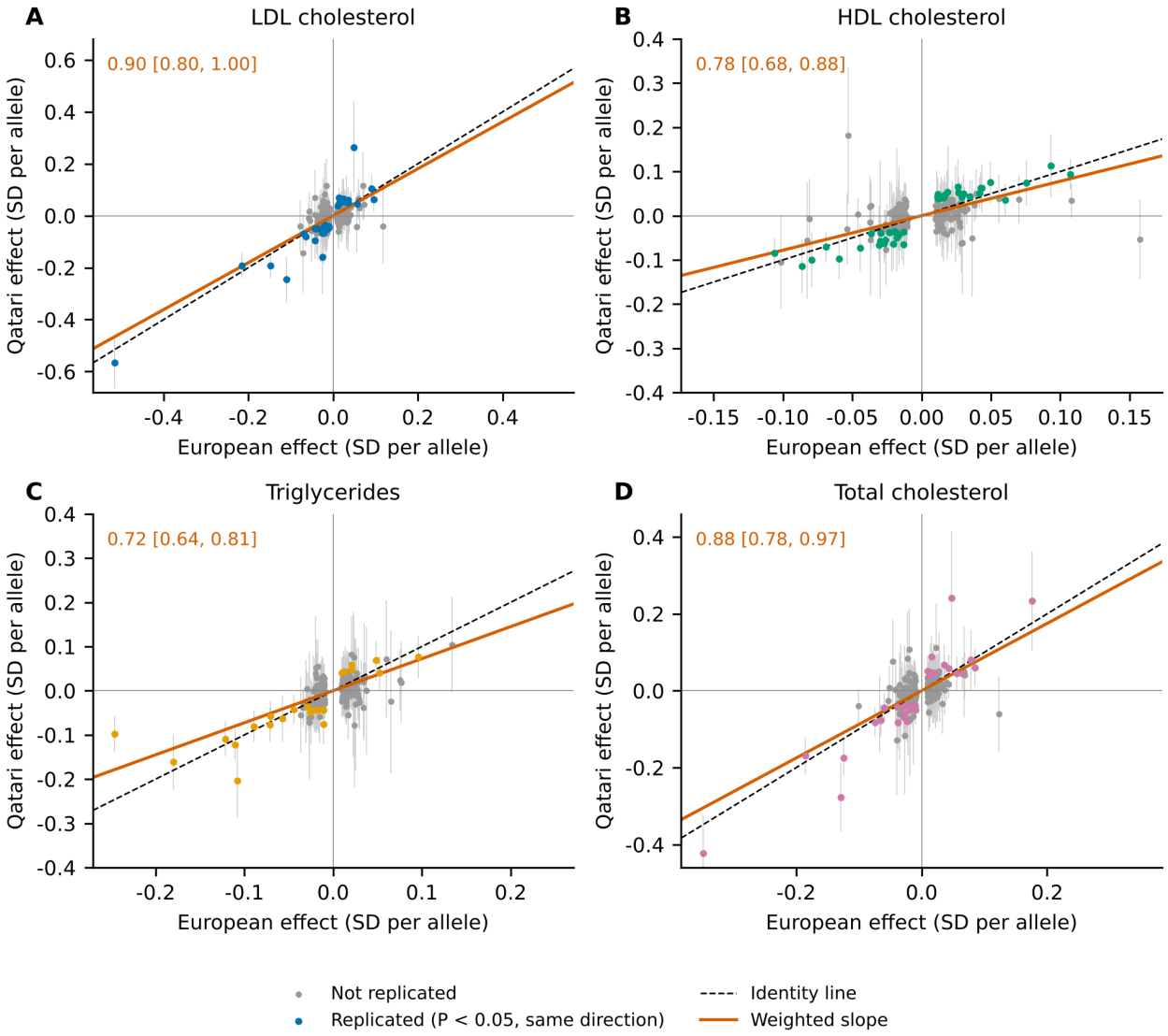


Figure 1. Per-allele effects of European-ancestry lipid index variants in European (x-axis) and Qatari (y-axis) summary statistics for (A) LDL cholesterol, (B) HDL cholesterol, (C) triglycerides and (D) total cholesterol. Coloured points replicated nominally ($P < 0.05$, same direction); grey points did not. Vertical lines are Qatari 95% confidence intervals. The dashed line is identity; the solid line is the inverse-variance-weighted slope, printed with its 95% confidence interval. Effects are in standard-deviation units of inverse-normal-transformed lipid levels per effect allele. LDL, low-density lipoprotein; HDL, high-density lipoprotein; SD, standard deviation.

3.3 Effect-size calibration and heterogeneity

Qatari effects were generally smaller than European effects (Figure 1), although the 95% CI for LDL-C reached 1. The weighted slope was 0.90 (95% CI 0.80–1.00) for LDL-C, 0.88 (0.78–0.97) for TC, 0.78 (0.68–0.88) for HDL-C and 0.72 (0.64–0.81) for TG. Effect-size correlations ranged from 0.53 (HDL-C) to 0.77 (LDL-C). The slopes were robust in sensitivity analyses. Restricting to variants with Qatari MAF ≥ 0.05 gave slopes

of 0.84 (LDL-C), 0.82 (HDL-C), 0.71 (TG) and 0.82 (TC). Restricting to strongly associated European variants ($P < 5 \times 10^{-12}$) gave 0.92, 0.81, 0.74 and 0.88, and the frequency-based SE gave 0.90, 0.78, 0.73 and 0.88. Nominal heterogeneity ($P < 0.05$) was observed for 6.6% to 7.8% of variants, close to the 5% expected by chance. Only two variant-trait pairs remained heterogeneous after Bonferroni correction. For TG, rs964184 in the *APOA5–ZPR1* region had a per-allele effect of −0.098 SD in Qatari data, against −0.246 SD in European data. For HDL-C, rs538969346 near *SMAD7* is rare in Europeans (effect-allele frequency 0.010) but more common in Qataris (0.034), and its effect was in the opposite direction. Attenuation was therefore diffuse and modest at individual loci rather than driven by a few strongly discordant loci.

Post hoc distance-pruning sensitivity at 1 Mb and 5 Mb was added during editorial review. At 5 Mb, slopes (95% CI) were LDL-C 0.949 (0.841–1.056), HDL-C 0.823 (0.711–0.934), TG 0.738 (0.645–0.830) and TC 0.944 (0.838–1.050). Distance pruning does not establish LD independence. Poisson-binomial probabilities and original intervals remain conditional on independent tests and fixed European effects; they may be over-precise. Wider pruning weakens the evidence for TC attenuation. These are sensitivity analyses, not prespecified results.

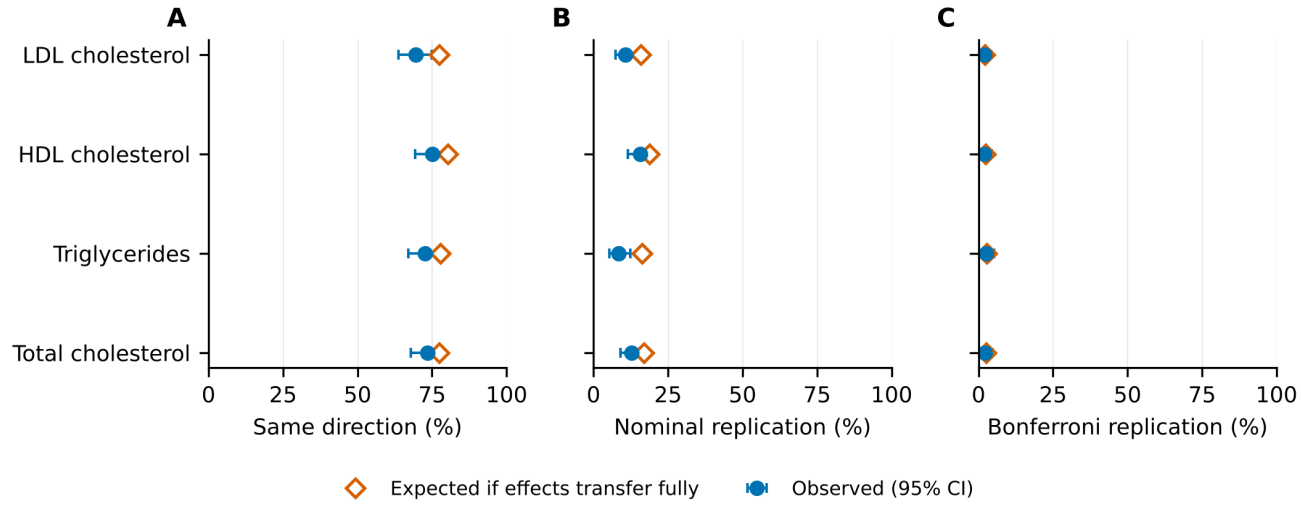


Figure 2. Observed (circles with 95% confidence intervals) and expected (open diamonds, assuming full transfer of European effects) percentages of index variants with (A) the same direction of effect, (B) nominal replication ($P < 0.05$, same direction) and (C) Bonferroni replication in the Qatari summary statistics. Expected percentages are the mean, across variants, of the probability of each event given the European effect and the Qatari standard error.

3.4 Polygenic score transfer

Pruning removed one HDL-C variant, so the score analyses used 269, 260, 274 and 283 variants. Under full transfer, the European-weighted score would explain 7.91% of LDL-C variance at Qatari allele frequencies, compared with 9.81% at European frequencies. Allele-frequency differences alone therefore reduce the expected variance explained to 0.81 of the European value for LDL-C and to 0.89 for TC. For HDL-C (1.01) and TG (1.07) these differences had no material effect (Table 2, Figure 3B). The observed variance explained in Qatari data was 6.46% (95% CI 5.25–7.82) for LDL-C, 5.63% (4.51–6.92) for TC, 4.13% (3.23–5.17) for TG and 3.57% (2.75–4.52) for HDL-C. Relative to expectation at Qatari frequencies, this corresponds to relative R^2 values of 0.82 (0.66–0.99), 0.77 (0.61–0.94), 0.53 (0.41–0.66) and 0.60 (0.47–0.77), respectively. The score association was highly significant for every trait (all $P < 0.001$). Its per-unit effect equalled the calibration slope (for example, 0.90, 95% CI 0.81–0.99, for LDL-C).

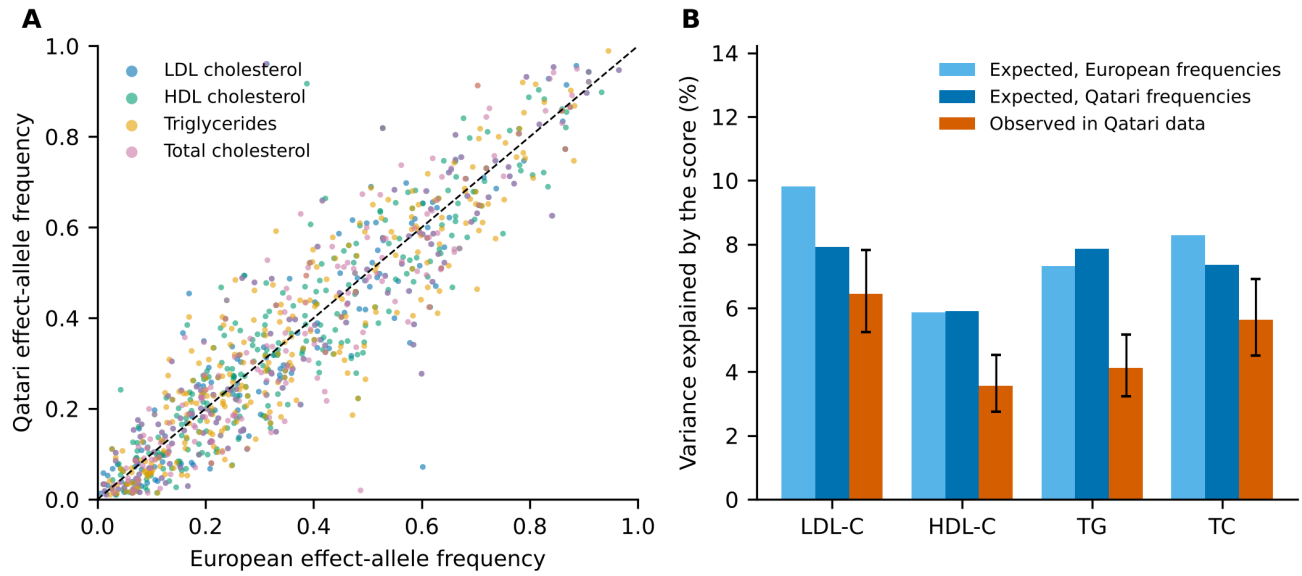


Figure 3. (A) Effect-allele frequencies of analysed index variants in European and Qatari data, by trait; the dashed line is identity. (B) Variance in each lipid trait explained by a European-weighted score of the index variants: expected under full transfer at European and at Qatari allele frequencies, and observed in Qatari summary statistics with 95% confidence intervals. LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglycerides; TC, total cholesterol.

3.5 Qatari genome-wide significant loci and fine-mapping

Distance clumping yielded 22 Qatari genome-wide significant loci: 5 for LDL-C, 5 for HDL-C, 7 for TG and 5 for TC. All 22 lay within 500 kb of a same-trait European index variant, so none was ancestry-specific at this resolution. Twenty-one lay within 250 kb and were fine-mapped. Five of these, all on chromosome 11q23 (116.5–118.0 Mb) for HDL-C and TG, were multi-signal regions. Their leads were low-frequency variants (effect-allele frequency 0.011 to 0.014) with large effects, spread over about 1.3 Mb. This pattern is consistent with one extended haplotype carried by a minority of Qatari participants rather than five independent loci. This interpretation could not be tested without an Arab LD reference.

Among the 16 single-signal loci, the median 95% credible set contained 6 variants (range 1 to 53), and 7 sets contained no more than 5 variants (Table 3, Figure 4). Three sets resolved to a single variant: *APOE* rs7412 (the ε2-defining allele) for LDL-C and for TC, and *GCKR* rs1260326 for TG. The European index variant was present in the Qatari data at 15 of these loci and fell inside the Qatari 95% credible set at 12. At *CELSR2*, the Qatari credible set for LDL-C contained the functional *SORT1* regulatory variant rs12740374 [32]. At the *TM6SF2* locus, the Qatari signal peaked at rs111234557, about 57 kb from the European index missense variant rs58542926 [33]. The European index variant lay outside the Qatari credible set at *APOB* (LDL-C and TC) and *LPL* (HDL-C). Credible-set sizes changed little with prior SDs of 0.10 or 0.20 (Supplementary data).

Table 3. Single-causal-variant Bayesian fine-mapping of genome-wide significant Qatari loci that overlap European index variants.

Trait	Locus (gene)	Chr	Variants	Top variant (PIP)	95% credible set size	European index variant	Index PIP	Index in set
LDL-C	*APOE*	19	1,622	rs7412 (1.00)	1	rs7412	1.00	Yes
LDL-C	*LDLR*	19	1,252	rs112736558 (0.06)	28	rs73015024	0.04	Yes
LDL-C	*CELSR2*	1	1,268	rs12740374 (0.23)	6	rs12740374	0.23	Yes

Trait	Locus (gene)	Chr	Variants	Top variant (PIP)	95% credible set size	European index variant	Index PIP	Index in set
LDL-C	*APOB*	2	1,325	rs1800481 (0.33)	9	rs934197	0.00	No
LDL-C	*TM6SF2*	19	1,230	rs111234557 (0.67)	8	rs58542926	0.10	Yes
HDL-C	*CETP*	16	1,560	rs708272 (0.56)	2	rs183130	NA	Not in Qatari data
HDL-C	*LPL*	8	2,137	rs74869266 (0.36)	27	rs15285	0.00	No
HDL-C	*ALDH1A2*	15	1,842	rs1077834 (0.80)	6	rs261290	0.05	Yes
TG	*GCKR*	2	719	rs1260326 (1.00)	1	rs1260326	1.00	Yes
TG	*MLXIPL*	7	1,032	rs33951980 (0.13)	53	rs3812316	0.05	Yes
TG	*TM6SF2*	19	1,230	rs111234557 (0.83)	3	rs58542926	0.03	Yes
TC	*APOE*	19	1,622	rs7412 (1.00)	1	rs7412	1.00	Yes
TC	*CELSR2*	1	1,269	rs7528419 (0.38)	5	rs12740374	0.24	Yes
TC	*LDLR*	19	1,272	rs8106503 (0.11)	27	rs73015024	0.04	Yes
TC	*TM6SF2*	19	1,230	rs111234557 (0.92)	2	rs58542926	0.03	Yes
TC	*APOB*	2	1,325	rs1800481 (0.20)	36	rs934197	0.00	No

Chr, chromosome; PIP, posterior inclusion probability from Wakefield approximate Bayes factors (prior standard deviation 0.15) within ±250 kb of the Qatari lead variant; NA, not applicable. Five loci on chromosome 11q23 with several genome-wide significant signals within 1 Mb were excluded because the single-causal-variant assumption is not tenable there. Abbreviations as in Table 1.

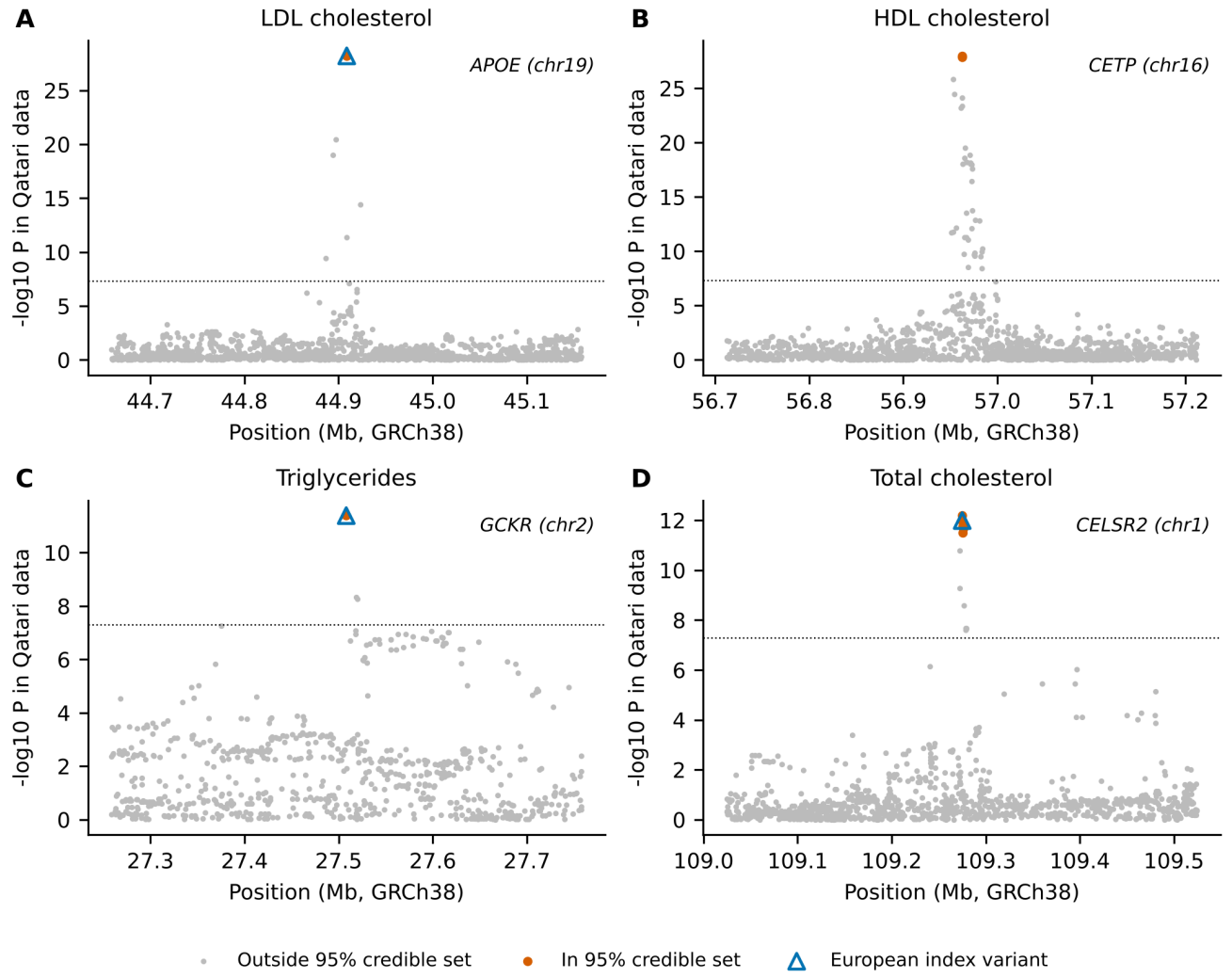


Figure 4. Regional association and fine-mapping in Qatari summary statistics at (A) APOE for LDL cholesterol, (B) CETP for HDL cholesterol, (C) GCKR for triglycerides and (D) CELSR2 for total cholesterol. Orange points form the 95% credible set; the open triangle marks the European index variant; the dotted line marks P = 5 × 10−8. Positions are GRCh38. Credible sets assume a single causal variant per region (Wakefield approximate Bayes factors, prior standard deviation 0.15). The European index variant at CETP (rs183130) is absent from the Qatari data.

4. Discussion

4.1 Principal findings

In 1,087 variant-trait pairs, almost three quarters had concordant effect directions. Large-effect variants replicated at approximately the expected rate, while nominal replication was lower than the full-transfer expectation. These comparisons are conditional on European point estimates, independent variant tests and comparable phenotype scales. They do not establish a biological cause of attenuation or separate true effect differences from linkage disequilibrium, selection, measurement and winner's-curse effects.

4.2 Comparison with previous studies

Hebbar and colleagues concluded that GWAS-identified metabolic variants showed low transferability to the Arabian Peninsula [18]. Our power calibration qualifies that conclusion. With a median power of about 10% per variant, a Qatari cohort of 6,000 would replicate only a small minority of European variants even if every effect transferred perfectly. Counting significant replications therefore understates transferability, and the appropriate comparison is with the expected count. Ismail Umlai and colleagues reported a moderate correlation of LDL-C effect sizes between Qatari and GLGC data using individual-level Qatar Biobank data [19]. Our open-data cal-

ibration slope for LDL-C (0.90) and effect correlation (0.77) agree in direction, and we extend the comparison to HDL-C, TG and TC. Shim and colleagues showed that polygenic scores for cardiometabolic disease, once optimised for Arabs using public data, performed better than published scores and comparably to European-ancestry benchmarks in Saudi Arab patients [20]. Our relative R^2 estimates quantify the loss that occurs when European weights are applied without such optimisation, at the level of individual lipid traits, using data that anyone can reproduce.

Cross-ancestry studies support shared genetic signals and reduced portability of scores across ancestry groups [9–11,29–31]. The lower slopes observed here are compatible with differences in linkage disequilibrium, phenotype definition, selection and environment. This design cannot identify which mechanism accounts for the differences or demonstrate that causal effects are attenuated.

The fine-mapping results add evidence on mechanism. The *CELSR2–SORT1* credible set contained rs12740374, the regulatory variant shown experimentally to alter hepatic *SORT1* expression [32]. The *TM6SF2* signal in Qatari data peaked near, but not at, the known E167K missense variant [33], which justifies follow-up with Qatari LD information. At *APOB* and *LPL*, the European index variants were outside the Qatari credible sets. These are candidate loci at which Arab LD structure may help distinguish causal from tagging variants, which is the main argument for diverse fine-mapping made by GLGC [8]. The clustered low-frequency signals on chromosome 11q23, near the *APOA1–APOC3–APOA4–APOA5* cluster, are consistent with an extended haplotype, as expected in a population with long runs of homozygosity [13,15]. They also overlap regions where familial and population-enriched lipid variants have been described in Qatar [34]. The strongest locus-level heterogeneity in the study was also in this region: the European index variant rs964184 had less than half its European effect on triglycerides in Qatari data. This is consistent with the view that the causal architecture of the *APOA5* region differs between the two populations, possibly because a Qatari low-frequency haplotype carries part of the signal that common variants tag in Europeans.

Several mechanisms could contribute to lower tag-variant slopes, including differences in linkage disequilibrium, gene–environment relationships, treatment adjustment and phenotype measurement. European discovery effects are also estimated with error and selected for significance. Restriction to strongly associated variants does not eliminate winner's curse. The estimates should therefore be interpreted as descriptive transfer of published tag-variant associations, rather than estimates of ancestry-specific causal effects.

The power-calibrated framework also has methodological value beyond lipids. Replication studies in under-represented populations are often small, and binary declarations of replication or non-replication are easily misread as evidence of absent transferability. Comparing observed with expected counts, and estimating a calibration slope with its CI, turns the same summary statistics into quantitative estimates of transfer. Because the framework needs only effect sizes, SEs and allele frequencies, it can be applied to any trait for which Arab summary statistics are released, including glycaemic traits, blood pressure and vitamin D concentrations.

4.3 Implications for Saudi Arabia and the Gulf

Saudi dyslipidaemia guidance emphasises risk detection [3], and polygenic approaches are being studied in Arabs [20]. The present summary-statistic analysis does not establish clinical utility or absolute cardiovascular risk. Multiplying all weights by a common slope changes the score scale but does not improve its ranking, correlation or explained variance. Clinical translation requires an Arab linkage-disequilibrium reference, appropriate phenotype harmonisation and independent individual-level validation. Results in Qatar cannot be assumed to apply directly to Saudi populations [13,14].

4.4 Strengths and limitations

The study has several strengths. It relies entirely on public, checksum-verified summary statistics and a single reproducible script. It compares replication with explicit power-based expectations rather than counting significant results. It separates the effect of allele-frequency differences from the effect of attenuated effect sizes on score performance. It covers all four standard lipid traits, and its effect-size calibration was stable across sensitivity analyses.

There are also limitations. First, we had no Arab LD reference, so fine-mapping assumed a single causal variant per region, and multi-signal regions were excluded rather than resolved. Second, the R^2 estimates assume independent index variants. Some residual LD among GLGC index variants within 500 kb windows remains possible, although pruning removed only one variant. Third, Qatari SEs were reconstructed from P values. Agreement with the frequency-based approximation was very high, but rounding of very small P values could add minor error. Fourth, about one third of European index variants were unavailable in the Qatari data, mostly because they were rare or absent. Transferability of rare variants therefore could not be assessed. Fifth, European effects can carry residual winner's curse, which would bias slopes downwards and inflate the expected replication counts. The slope restricted to strongly associated variants was similar, however, which argues against a large bias. Sixth, both datasets measured lipids without full harmonisation of fasting status or treatment adjustment, and GLGC adjusted LDL-C and TC for lipid-lowering medication. Such differences could contribute to attenuation. Seventh, Qatari participants were used as a proxy for Middle Eastern Arabs; transfer to Saudi and other Gulf Arab populations was not tested directly. Finally, the Qatari sample is modest, so the CIs for trait-specific estimates are wide.

Inferential scope: R^2 estimates, intervals and replication probabilities assume variant independence and fixed European effects. Distance pruning does not establish LD independence. Original estimates are therefore conditional exploratory summaries; nominal significance must not be used to establish ancestry-specific causal differences or clinical utility.

4.5 Future research

Future work should combine Arab LD panels with multi-ancestry fine-mapping and colocalisation [40] to resolve the *APOB*, *LPL*, *TM6SF2* and chromosome 11q23 signals. Locally calibrated and multi-ancestry polygenic scores should be validated against incident cardiovascular events in Saudi and Gulf cohorts [36,38]. The analysis should be extended to protein and metabolite traits, for which differences between Arab and European genetic architecture have been reported [39].

5. Conclusions

Published European lipid associations showed substantial directional concordance in Qatari Arabs. HDL-C and TG slopes remained below 1 under wider distance pruning, while LDL-C and TC inference was more sensitive. Summary-based score variance and single-signal credible sets remain conditional exploratory estimates. Individual-level validation and Arab linkage-disequilibrium information are required before clinical application.

Declarations

Funding: NA

Conflicts of Interest: NA

Institutional Review Board Statement: This study used only publicly available, de-identified aggregate genome-wide association summary statistics from the NHGRI-EBI GWAS Catalog; no individual-level data were accessed and no ethics approval is claimed.

Informed Consent Statement: Not applicable (public de-identified aggregate data).

Data Availability Statement: The Qatar Genome Programme summary statistics are available from the NHGRI-EBI GWAS Catalog (accessions GCST90013331, GCST90013332, GCST90013333 and GCST90013334; https://www.ebi.ac.uk/gwas/), and the GLGC 2021 European-ancestry associations are available from the same resource (accessions GCST90239652, GCST90239658, GCST90239664 and GCST90239676). The analysis code and derived variant-level results are available with the study materials.

Author Contributions: NA

References

[1] Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. J Am Coll Cardiol. 2020;76(25):2982-3021. doi:10.1016/j.jacc.2020.11.010

[2] Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J. 2017;38(32):2459-2472. doi:10.1093/eurheartj/ehx144

[3] AlRahimi J, AlSaif S, Alasnag M, Awan Z, Almutairi F, Al Mudaiheem H, et al. 2022 Saudi guidelines for the management of dyslipidemia. Heart Views. 2023;24(2):67-92. doi:10.4103/heartviews.heartviews_102_22

[4] AlMuhaidib S, AlBuhairan F, Tamimi W, AlDubayee M, AlAqeel A, Babiker A, et al. Prevalence and factors associated with dyslipidemia among adolescents in Saudi Arabia. Sci Rep. 2022;12:16888. doi:10.1038/s41598-022-21262-9

[5] Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161-164. doi:10.1038/538161a

[6] Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. Cell. 2019;177(1):26-31. doi:10.1016/j.cell.2019.02.048

[7] Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x

[8] Graham SE, Clarke SL, Wu KHH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-3

[9] Wang Y, Guo J, Ni G, Yang J, Visscher PM, Yengo L. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. Nat Commun. 2020;11:3865. doi:10.1038/s41467-020-17719-y

[10] Privé F, Aschard H, Carmi S, Folkersen L, Hoggart C, O'Reilly PF, et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. Am J Hum Genet. 2022;109(1):12-23. doi:10.1016/j.ajhg.2021.11.008

[11] Ding Y, Hou K, Xu Z, Pimplaskar A, Petter E, Boulier K, et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. Nature. 2023;618(7966):774-781. doi:10.1038/s41586-023-06079-4

[12] Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet. 2024;25(1):8-25. doi:10.1038/s41576-023-00637-2

[13] Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, Gohar A, et al. Population genetic structure of the people of Qatar. Am J Hum Genet. 2010;87(1):17-25. doi:10.1016/j.ajhg.2010.05.018

[14] Fakhro KA, Staudt MR, Ramstetter MD, Robay A, Malek JA, Badii R, et al. The Qatar genome: a population-specific tool for precision medicine in the Middle East. Hum Genome Var. 2016;3:16016. doi:10.1038/hgv.2016.16

[15] Mbarek H, Devadoss Gandhi G, Selvaraj S, Al-Muftah W, Badji R, Al-Sarraj Y, et al. Qatar genome: insights on genomics from the Middle East. Hum Mutat. 2022;43(4):499-510. doi:10.1002/humu.24336

[16] Thareja G, Al-Sarraj Y, Belkadi A, Almotawa M, Qatar Genome Program Research (QGPR) Consortium, Suhre K, et al. Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. Nat Commun. 2021;12:1250. doi:10.1038/s41467-021-21381-3

[17] Al Kuwari H, Al Thani A, Al Marri A, Al Kaabi A, Abderrahim H, Afifi N, et al. The Qatar Biobank: background and methods. BMC Public Health. 2015;15:1208. doi:10.1186/s12889-015-2522-7

[18] Hebbar P, Abu-Farha M, Abubaker J, Channanath AM, Al-Mulla F, Thanaraj TA. Generalizability of GWA-identified genetic risk variants for metabolic traits to populations from the Arabian Peninsula. Genes (Basel). 2021;12(10):1637. doi:10.3390/genes12101637

[19] Ismail Umlai UK, Toor SM, Al-Sarraj YA, Mohammed S, Al Hail MS, Ullah E, et al. A multi-ancestry genome-wide association study and evaluation of polygenic scores of LDL-C levels. J Lipid Res. 2025;66(3):100752. doi:10.1016/j.jlr.2025.100752

[20] Shim I, Kuwahara H, Chen N, Hashem MO, AlAbdi L, Abouelhoda M, et al. Clinical utility of polygenic scores for cardiometabolic disease in Arabs. Nat Commun. 2023;14:6535. doi:10.1038/s41467-023-41985-1

[21] Wakil SM, Ram R, Muiya NP, Mehta M, Andres E, Mazhar N, et al. A genome-wide association study reveals susceptibility loci for myocardial infarction/coronary artery disease in Saudi Arabs. Atherosclerosis. 2016;245:62-70. doi:10.1016/j.atherosclerosis.2015.11.019

[22] Ram R, Wakil S, Muiya N, Andres E, Mazhar N, Hagos S, et al. A common variant association study in ethnic Saudi Arabs reveals novel susceptibility loci for hypertriglyceridemia. Clin Genet. 2017;91(3):371-378. doi:10.1111/cge.12859

[23] Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med. 2009;6(2):e1000022. doi:10.1371/journal.pmed.1000022

[24] Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51(D1):D977-D985. doi:10.1093/nar/gkac1010

[25] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-665. doi:10.1002/gepi.21758

[26] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-560. doi:10.1136/bmj.327.7414.557

[27] Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol. 2009;33(1):79-86. doi:10.1002/gepi.20359

[28] Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet. 2012;44(12):1294-1301. doi:10.1038/ng.2435

[29] Marigorta UM, Navarro A. High trans-ethnic replicability of GWAS results implies common causal variants. PLoS Genet. 2013;9(6):e1003566. doi:10.1371/journal.pgen.1003566

[30] Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514-518. doi:10.1038/s41586-019-1310-4

[31] Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet. 2018;50(3):390-400. doi:10.1038/s41588-018-0047-6

[32] Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010;466(7307):714-719. doi:10.1038/nature09266

[33] Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2014;46(4):352-356. doi:10.1038/ng.2901

[34] Diboun I, Al-Sarraj Y, Toor SM, Mohammed S, Qureshi N, Al Hail MSH, et al. The prevalence and genetic spectrum of familial hypercholesterolemia in Qatar based on whole genome sequencing of 14,000 subjects. Front Genet. 2022;13:927504. doi:10.3389/fgene.2022.927504

[35] Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10:3328. doi:10.1038/s41467-019-11112-0

[36] Ruan Y, Lin YF, Feng YCA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nat Genet. 2022;54(5):573-580. doi:10.1038/s41588-022-01054-7

[37] Kanoni S, Graham SE, Wang Y, Surakka I, Ramdas S, Zhu X, et al. Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis. Genome Biol. 2022;23:268. doi:10.1186/s13059-022-02837-1

[38] Urbut SM, Chen Q, Sui Y, Misra A, Koyama S, Hornsby W, et al. Genetics of cholesterol and coronary disease risk across six global ancestries. NEJM Evid. 2025;4(11):EVIDoa2500105. doi:10.1056/EVIDoa2500105

[39] Thareja G, Belkadi A, Arnold M, Albagha OME, Graumann J, Schmidt F, et al. Differences and commonalities in the genetic architecture of protein quantitative trait loci in European and Arab populations. Hum Mol Genet. 2023;32(6):907-916. doi:10.1093/hmg/ddac243

[40] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383

Published by Lumora Journals · lumora.sa · CC BY 4.0