

A multi-ancestry genome-wide association study and evaluation of polygenic scores of LDL-C levels

Umm-Kulthum Ismail Umlai¹, Salman M. Toor¹, Yasser A. Al-Sarraj^{1,2}, Shaban Mohammed³, Moza S. H. Al Hail³, Ehsan Ullah⁴, Khalid Kunji⁴, Ayman El-Menyar⁵, Mohammed Gomaa⁶, Amin Jayyousi⁷, Mohamad Saad⁴, Nadeem Qureshi⁸, Jassim M. Al Suwaidi⁶, and Omar M. E. Albagha^{1,9,*}

¹College of Health and Life Sciences (CHLS), Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), Doha, Qatar; ²Qatar Genome Program (QGP), Qatar Foundation Research, Development and Innovation, Qatar Foundation, Doha, Qatar; ³Department of Pharmacy, Hamad Medical Corporation, Doha, Qatar; ⁴Qatar Computing Research Institute (QCRI), Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), Doha, Qatar; ⁵Trauma and Vascular Surgery, ⁶Adult Cardiology, Heart Hospital, and ⁷Department of Diabetes, Hamad Medical Corporation (HMC), Doha, Qatar; ⁸Primary Care Stratified Medicine Research Group, Centre for Academic Primary Care, School of Medicine, University of Nottingham, Nottingham, United Kingdom; and ⁹Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom

Abstract The genetic determinants of low-density lipoprotein cholesterol (LDL-C) levels in blood have been predominantly explored in European populations and remain poorly understood in Middle Eastern populations. We investigated the genetic architecture of LDL-C variation in Qatar by conducting a genome-wide association study (GWAS) on serum LDL-C levels using whole genome sequencing data of 13,701 individuals (discovery; $n = 5,939$, replication; $n = 7,762$) from the population-based Qatar Biobank (QBB) cohort. We replicated 168 previously reported loci from the largest LDL-C GWAS by the Global Lipids Genetics Consortium (GLGC), with high correlation in allele frequencies ($R^2 = 0.77$) and moderate correlation in effect sizes (Beta; $R^2 = 0.53$). We also performed a multi-ancestry meta-analysis with the GLGC study using MR-MEGA (Meta-Regression of Multi-Ethnic Genetic Association) and identified one novel LDL-C-associated locus; *rs10939663* (*SLC2A9*; genomic control-corrected $P = 1.25 \times 10^{-8}$). Lastly, we developed Qatari-specific polygenic score (PGS) panels and tested their performance against PGS derived from other ancestries. The multi-ancestry-derived PGS (PGS000888) performed best at predicting LDL-C levels, whilst the Qatari-derived PGS showed comparable performance. Overall, we report a novel gene associated with LDL-C levels, which may be explored further to decipher its potential role in the etiopathogenesis of cardiovascular diseases. Our findings also highlight the importance of population-based genetics in developing PGS for clinical utilization.

Supplementary key words genomics • LDL • cholesterol • dyslipidemias • atherosclerosis

Cardiovascular diseases (CVDs) remain the leading cause of global mortality. Elevated low-density lipoprotein cholesterol (LDL-C) concentration in blood, which is a known heritable and modifiable risk factor for the development of atherosclerotic CVD, presents as an increasingly prevalent global health burden (1, 2). Notably, the incidence of CVD is on the rise in the Middle Eastern region (3). The incidence of death per 100,000 individuals attributed to CVD is also higher in the Middle East compared to Europe and the United States and predominantly affects males aged 58–59 years, which is a younger age band compared to the rest of the world (4).

An elevated level of LDL-C is identified as one of the main risk factors for the development of CVD, and lowering LDL-C concentration leads to a significant reduction in major vascular and coronary complications and stroke (5, 6). Deciphering the genetic determinants of variation in LDL-C levels has merits of improving the understanding of LDL-C regulation. Additionally, it has the potential for identifying novel molecular targets and for resolving the paradigm around the development and progression of atherosclerotic CVD. Pathogenic variants in genes involved in LDL-C uptake and catabolism such as *LDLR*, *APOB*, *APOE*, *LDLRAP1*, and *PCSK9* are recognized as the principal causative factors for monogenic familial hypercholesterolemia (FH), one of the most common inherited autosomal condition in the world (7). A recent study showed that the prevalence of FH was estimated at 1 in 125 individuals in the Middle Eastern population of Qatar, which is higher than the global prevalence (8).

*For correspondence: Omar M. E. Albagha, oalbagha@hbku.edu.qa.

However, variation in LDL-C concentration is a quantitative trait that is broadly dictated by a polygenic cause due to common alleles with small/moderate effects, which influence multiple LDL-C modulating genes that can lead to raised LDL-C levels.

Genome-wide association studies (GWAS) have enabled the identification of several genetic risk loci that are associated with blood lipid concentrations (9). The GWAS catalog currently lists over 1,850 genomic risk loci associated with blood lipids (10). Notably, a recent comprehensive GWAS conducted by the Global Lipids Genetics Consortium (GLGC), which represents the largest published GWAS and meta-analysis on LDL-C levels (Graham *et al.*) was performed on a cohort of around 1.65 million individuals (11). The study is an aggregate of 201 studies from 5 ancestries: African (6.0%), East Asian (8.9%), European (79.8%), South Asian (2.5%), and Hispanic (2.9%), with little representation of Middle Eastern populations. This large-scale multi-ancestry meta-analysis revealed more than 900 lipid-related genetic loci (11).

The inclusion of additional underrepresented ancestries, such as those of Arab descent, can improve the understanding of lipid-level genetic determinants, reveal novel drug targets, advance fine mapping of variants associated with lipid levels, and progress polygenic disease prediction (11). Associations of genetic variants with lipid traits and increased susceptibility to CVD are recurrently explored for the development of polygenic scores (PGS) (12–14). The earliest LDL-C PGS developed is a 12-SNP scoring panel from the GLGC, in 2013 (15). This scoring panel was later refined in another study by the same group to a 6-SNP score which performed equally well at discriminating between hypercholesterolemia patients with no confirmed mutation and healthy controls (16). This score has not been published on the Polygenic Score Catalog (PGS catalog) (17) but the PGS panel is available in the publication (16).

However, since GWAS has been predominantly conducted on individuals of European ancestry (18), the portability of genetic risk loci among different ancestries requires thorough investigations. This is due to differences in genetic architectures encompassing differences in allele frequencies and linkage disequilibrium (LD) that can affect genetic associations (19). For instance, genetic risk scores for lipid traits developed in European populations did not perform well in Sub-Saharan African populations (20, 21). Population-based GWAS of underrepresented regions can reveal both common and rare genetic variants related to the inherited risk of elevated LDL-C concentration and increased susceptibility to CVD, while multi-ancestry meta-analyses may also be advantageous for uncovering novel variants. Increasing diversity in LDL-C GWAS as opposed to mitigating the limitations of currently devised PGS has merits for improving the overall utilization of PGS prediction.

PGS development for predicting LDL-C levels or CVD within the wider Middle Eastern region based on local genomic variation has not been systematically explored. A recent study on a cohort of coronary heart disease (CHD) from Qatar showed that the performance of PGS derived from European studies of CVD was comparable to the performance reported in Europeans (22). However, European-derived polygenic scores for many quantitative traits performed lower when assessed in the Qatari population compared to European populations. This was revealed in a GWAS conducted in Qatar, which focused on 45 clinically relevant traits (23).

In this study, we conducted a comprehensive GWAS of LDL-C in the Qatari population based on whole genome sequencing (WGS) data of 13,701 individuals from the Qatar Biobank (QBB) cohort. We also performed a multi-ancestry meta-analysis of our GWAS data with the largest published data from the GLGC consortium to broaden the spectrum of LDL-C-associated loci. Importantly, we derived Qatari-specific PGS panels and assessed their performance to predict LDL-C levels against PGSs derived from African, East Asian, European, South Asian, and multi-ancestry populations. Overall, our findings contribute towards diversifying the current knowledge of population genetics for LDL-C levels and aid the clinical application of PGS. This study represents the largest GWAS of Arab populations performed on LDL-C association and polygenic risk score panels derived from the largest discovery and replication cohort within Arab populations.

MATERIALS AND METHODS

Participant recruitment and data collection

Study participants were recruited from the Qatar Biobank (QBB), a national biorepository, which collects biological samples and clinicopathological data from Qatari citizens and long-term residents, in coordination with the national public healthcare provider (Hamad Medical Corporation). Our study cohort comprised 14,409 subjects aged 18–89 years, of whom 14,156 (98.24%) were Qatari nationals and 253 (1.76%) were long-term residents. The present study focused on Qatari nationals and the first data release of QBB was used as a discovery dataset and the replication dataset was based on the second release. All study participants provided written informed consent prior to inclusion in this study. This study was approved by the institutional review boards of QBB (Protocol no. E-2020-QF-QBB-RES-ACC-0154-0133) and Hamad Bin Khalifa University (HBKU), Doha, Qatar (Approval no. QBRI-IRB 2021-03-078). The study was conducted according to the guidelines of the Declaration of Helsinki.

Biological samples (blood, urine, and saliva) were collected from study participants. In addition, routine medical examinations and biochemical tests (e.g. lipid profile, trace mineral concentrations, etc.) were also performed and clinical data were retrieved. Participants also filled out medical

questionnaires with questions pertaining to medical and family histories, medication, and supplements (24). Cholesterol treatment was determined based on self-reported data from questionnaires.

LDL-C measurement

LDL-C levels were calculated from blood lipid profiles (Total cholesterol, High-density lipoprotein-cholesterol, and Triglycerides) using the Friedewald formula (FF) and recorded in mg/dl. Serum lipid profiles were determined using enzymatic colorimetric assays performed using a Roche COBAS 6000 c analyzer (Roche, Switzerland).

Whole genome sequencing

Whole genome sequencing (WGS) was conducted by the Qatar Genome Program (QGP) (25), as previously described (23). Briefly, genomic DNA was isolated from whole blood using the QIAamp DNA Blood MIDI kit (Qiagen) and QIA-Symphony SP automated instrument (Qiagen) as per the manufacturer's instructions. DNA quality was evaluated by the Caliper Labchip GXII (PerkinElmer) Genomic DNA assay and concentrations were measured using the Quant-iT dsDNA Assay (Invitrogen). Libraries for WGS were prepared using the Illumina TruSeq DNA Nano kit (Illumina). The genomic libraries were sequenced using the HiSeq X Ten (Illumina) at 30X coverage at the Clinical Genomic Facilities at Sidra Medicine, Doha, Qatar.

Quality control and data preprocessing

FastQC (v0.11.2) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to perform quality control measures on the genomic data files and read alignment was performed against the GRCh38 genome reference, using bwa.kit (v0.7.12) (<https://github.com/lh3/bwa/tree/master/bwakit>). Picard (v1.117; [CollectWgsMetrics]) (<http://broadinstitute.github.io/picard/>) was used to control the quality of mapped reads and the variants were jointly called in a combined variant call file (gVCF) for the entire cohort following GATK 3.4 best practices (26). Downstream quality control and data preprocessing for GWAS analysis preparation were performed using the PLINK tool (v1.9) (27). Genetic variants with call rates < 90%, minor allele frequency (MAF) < 0.01, mean depth coverage < 10X, those with a Hardy-Weinberg equilibrium $P < 1.0 \times 10^{-6}$ were excluded from the analysis. Quality control measures were applied to whole genome sequencing data from 14,409 study subjects in which samples with genotype call rate < 95% ($n = 18$), excess heterozygosity ($n = 34$), duplicate samples ($n = 45$), gender ambiguity ($n = 251$), population outliers ($n = 253$; determined by multi-dimensional scaling) were excluded from the dataset. Additionally, subjects with missing LDL-C measurements ($n = 107$) were removed from the dataset. After quality control, the total number of subjects used in downstream analysis was 13,701 comprising 5,939 in the discovery set and 7,762 in the replication set.

GWAS analysis

LDL-C values were adjusted for age, age², the first ten genetic principal components (PC1 to PC10), and cholesterol treatment separately by gender, and residuals were calculated. GWAS was performed using SAIGE (v0.44.5) (28), applying inverse normalization of the LDL-C residuals and default

adjustment for relatedness in the data. To assess the replication of known loci, a list of variants from the largest published GWAS of LDL-C by Graham *et al.*, (11) was downloaded from the University of Michigan Center for Statistic Genetics (http://csg.sph.umich.edu/willer/public/glsc-lipids2021/results/trans_ancestry/). This list was filtered to only include variants meeting the genome-wide significance threshold ($p_{value_GC} < 5.0 \times 10^{-8}$) and then compared with the variants from the GWAS of the QBB discovery dataset with nominal significance ($P < 0.05$) to identify replicated variants. The R intersect function was used to compare data frames and output common SNP entries. The effect allele frequency and effect size (beta) from the multi-ancestry data of the Graham *et al.*, study was compared to that from our GWAS results using Spearman's correlation. Effect values and allele frequencies were then plotted using GraphPad PRISM (v9.0; GraphPad Software).

Multi-ancestry meta analysis

We first performed a fixed effect meta-analysis of QBB discovery and replication GWAS using the METAL tool (29). Then a multi-ancestry meta-analysis was performed using the results from METAL and with summary statistics from each of the single-ancestry LDL-C GWAS; African (AFR), East Asian (EAS), European (EUR), Hispanic (HIS), South Asian (SAS) from the GLGC study (11). Multi-ancestry meta-analysis was performed using the Meta-Regression of the Multi-AncEstry Genetic Association tool (MR-MEGA; v0.2) (30) using 3 PCs and genome control (GC) correction. Novel variants were identified based on attaining genome-wide significance (GC-corrected $P \leq 5.0 \times 10^{-8}$) post-meta-analysis, being beyond 500kb of any previously reported genome-wide significant SNPs from the GLGC study and not in LD with any previously reported genome-wide significant loci within 500 kb in the Common Metabolic Disease (CMD) Knowledge Portal (www.cmdkp.org) or in the GWAS Catalog (10). Beta effect values were calculated by running a multi-ancestry meta-analysis using METAL. Regional association plots were generated using the LocusZoom(v1.4) tool (31).

Analysis of polygenic scores

Polygenic score panels for LDL-C were all derived based on the QBB discovery dataset (training cohort) using three different methods (32); thresholding-only, clumping, and thresholding (C+T) or Bayesian shrinkage implemented in LDpred2 (33). For thresholding, eight different P -value thresholds were used, ranging from 5.0×10^{-4} to 5.0×10^{-16} . For C+T, GWAS results from the training set were pruned using the -clump option in PLINK (v1.9) at two linkage disequilibrium (LD) clumping r^2 cutoff values; 0.2 and 0.8, clumping distance of 500 kb and the same P -values cut off used in the thresholding method. Only index SNPs from the output of clumping were selected for the scoring panels. The effect size values were obtained from the beta-effect weights from the GWAS of the discovery cohort.

LDpred2 is a Bayesian shrinkage algorithm from the bigsnpr package on R (33), which was used to derive polygenic scores from the discovery cohort. The LDpred2 'grid' method was used for beta effect size estimation and prediction using the default parameters. A total of 168 different panels were tested by LDpred-2 and the 3 best performing panels were extracted for assessment in the QBB replication cohort.

PGS scoring for all panels were performed using “-score” and “sum” option in PLINK (v1.9). PGS panels obtained from other studies, derived from European (PGS000892, PGS000893 (11), PGS000814 (15), EUR_6SNP (Refined PGS000814) (16), PGS000875 (34)), Multi-ethnic (PGS000888, PGS000889), African (PGS000886, PGS000887), East Asian (PGS000891) and South Asian (PGS000897) (11) ancestries were downloaded from the Polygenic Score catalog (PGS catalog) or associated publication, and used to score the replication cohort.

The performance of PGS panels in predicting LDL-C was assessed using linear regression models with PGS, age, gender, cholesterol treatment, and the first four principal components (PCs) as predictors. Linear regression was performed using R software (v4.3.1) and the resulting adjusted R^2 values were used as performance metrics.

RESULTS

Characteristics of study cohort

An overview of the overall study design is depicted in Fig. 1. After quality control measures, the study cohort comprised 5,939 subjects in the discovery set and 7,762 subjects in the replication set (Total $n = 13,701$). Characteristic features of study cohorts are listed in supplemental Table S1.

LDL-C GWAS

Manhattan and QQ plots derived from the GWAS performed for the discovery dataset are shown in Fig. 2. The Manhattan plot illustrates 5 genome-wide significant loci located in chromosome bands 1p13.3, 2p24.1, 19q13.32, 19p13.2 and 19p13.11 (Fig. 2A). No widespread genomic inflation was observed as represented by the genomic inflation factor ($\lambda_{GC} = 1.05$), shown in the QQ-plot (Fig. 2B).

GWAS for the discovery dataset consisted of 8,399,151 variants, of which 5 loci (242 variants) reached genome-wide significance ($P < 5.0 \times 10^{-8}$). Meta-analysis of the discovery and replication cohort did not identify any novel LDL-C-associated variants (data not shown).

Next, we compared our findings from the GWAS of the discovery dataset with summary statistics of 86,607 genome-wide significant LDL-C-associated variants obtained from the Graham *et al.*, study (11) to assess the level of replication in our cohort. We replicated 8,043 variants (168 loci) at nominal significance ($P < 0.05$) in the QBB dataset (supplemental Table S2). Comparing the allele frequencies and effect sizes (Beta) of replicated variants to data from the Graham *et al.*, study showed a high correlation of allele frequencies

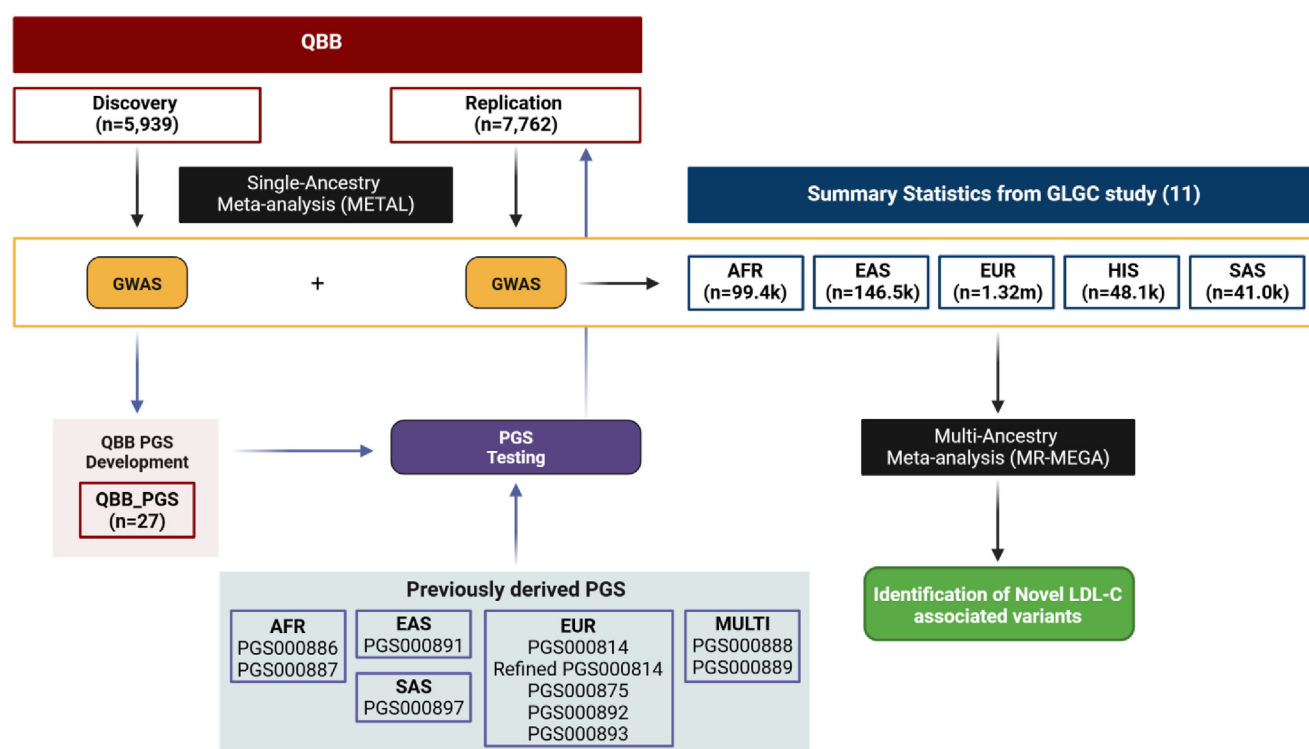


Fig. 1. Study design and analysis workflow. LDL-C GWAS was performed on the Qatar Biobank (QBB) discovery ($n = 5,939$) and replication ($n = 7,762$) cohorts separately using inverse-normalized LDL-C residuals. GWAS summary statistics from the QBB cohorts were then combined in a single-ancestry, fixed-effect meta-analysis using METAL. The results were further meta-analyzed using (MR-MEGA-v0.2) with 5 single-ancestry cohort summary statistics data (African; AFR; $n = 99.4k$, East-Asian: EAS; $n = 146.5k$, European; EUR; $n = 1.32m$, Hispanic; HIS; $n = 48.1k$ and South Asian; SAS; $n = 41.0k$) from the GLGC LDL-C study (11). GWAS results from the discovery cohort were then used to develop multiple PGS panels using different tools (refer to methods section). 27 QBB-derived PGS panels (QBB_PGS) were then tested on the QBB replication cohort. Previously derived PGS from other ancestries (African (AFR (11), European (EUR (11, 15, 16, 34)), East Asian (EAS (11)), South Asian (SAS (11)) and Multi-ancestry (MULTI (11)) were also tested in the QBB replication cohort.

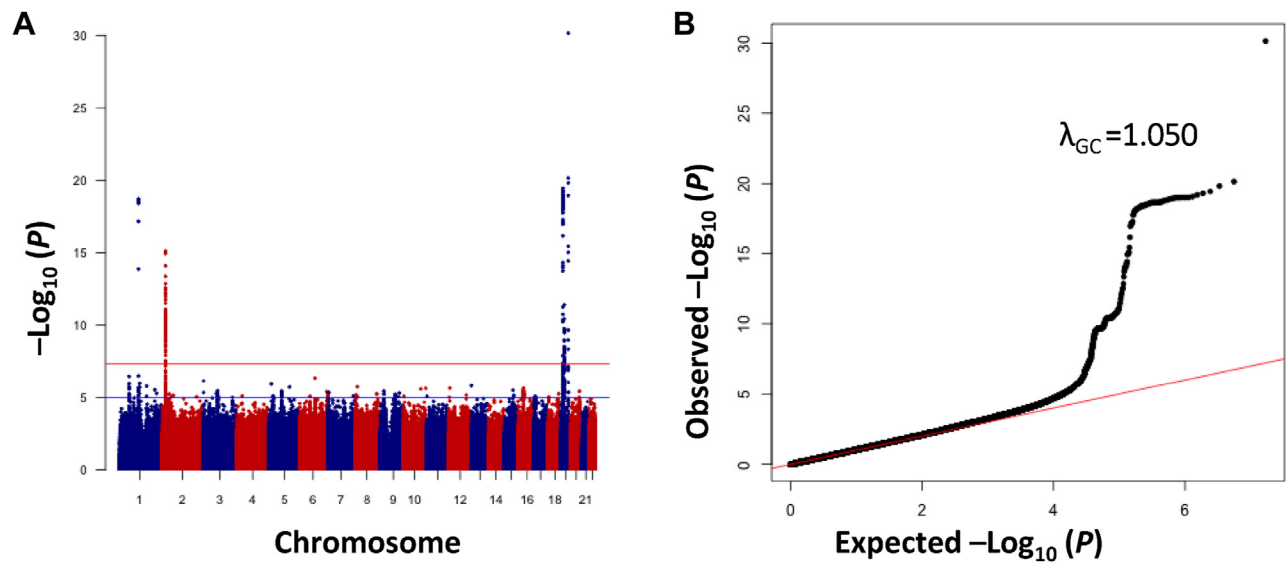


Fig. 2. Manhattan and Quantile-Quantile plots from QBB GWAS discovery cohort. A: The Manhattan plot represents genetic variants (dots) plotted on the x-axis in accordance with chromosome position against their corresponding $-\log_{10}(P)$. The horizontally marked red line indicates the genome-wide significance threshold ($P = 5.0 \times 10^{-8}$). The horizontal blue line indicates the suggestive significance threshold ($P = 1.0 \times 10^{-5}$). B: The quantile-quantile plot shows the expected versus observed $-\log_{10}(P)$; λ_{GC} is the genomic inflation factor. Analysis was based on data from the discovery cohort ($n = 5,939$) and performed using SAIGE adjusted for age, age², the first 10 genetic principal components (PC1 to PC10) applying inverse-normalization of the LDL-C residuals and adjustment for relatedness.

between the two datasets ($R^2 = 0.77$, Fig. 3A) and an overall consistent direction of effect, showing a significant trend for higher effect sizes in QBB cohort (Regression slope = 0.375; $P < 0.0001$; $R^2 = 0.53$, Fig. 3B).

Multi-ancestry meta-analysis of LDL-C

Next, we performed an MR-MEGA multi-ancestry meta-analysis using the QBB cohort alongside the summary statistics from each single-ancestry GWAS (AFR, EAS, EUR, HIS, SAS) from the Graham *et al.* study (11). The multi-ancestry meta-analysis yielded

52,621 variants with GC-corrected genome-wide significance. Notably, 2,472 of these variants (259 loci) only attained genome-wide significance after the meta-analysis. However, only one of these loci was novel, with no previously reported proxy SNPs in association with LDL-C reported in other studies/databases. The lead variant of our novel locus was rs10939663 (GC corrected $P = 1.25 \times 10^{-8}$) from the 4p16.1 locus, mapped within the Solute carrier family 2 member 9 *SLC2A9* (Table 1). The regional association plot for the novel locus is shown in Fig. 4.

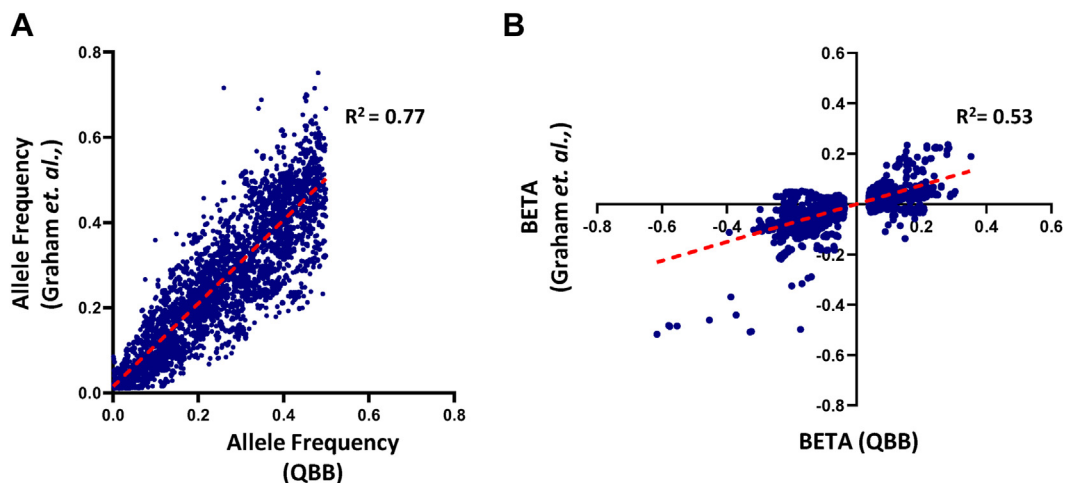


Fig. 3. Comparison of allele frequency and effect size for replicated genetic variants. Scatter plots represent the (A) Allele frequency and (B) Effect size (Beta) comparisons between the QBB discovery cohort ($n = 5,939$) and data from the multi-ancestry GWAS study by Graham *et al.*, ($n = 1,654,960$). R^2 is the coefficient of determination from correlation analysis. The red dotted line represents the line of best fit from linear regression.

TABLE 1. Lead genetic variant from a novel locus identified from multi-ancestry meta-analysis of QBB cohorts with the multi-ancestry summary statistics from Graham *et al.* (11)

CHR	BP	Locus	Nearest Gene	Variant	AF _{QBB}	AF _{IKGP}	A1	A2	<i>P</i> .value _{association}	Beta _{METAL}	<i>P</i> .value _{ancestry_het}	<i>P</i> .value _{residual_het}
4	10,030,892	chr4p16.1	SLC2A9	rs10939663	0.746	0.724	G	T	1.25×10^{-08}	-0.0072	5.04×10^{-05}	0.92

A1, Other allele; A2, Effect allele; AF_{IKGP}, Allele Frequency from 1000 Genomes Project (UK)- European subpopulation; AF_{QBB}, Allele Frequency from QBB discovery dataset; Beta_{METAL}, effect size value from fixed-effect METAL meta-analysis; BP, Base pairs; CHR, chromosome.

Derivation of Qatari-specific PGS panels

The clumping and thresholding methods were used to derive several PGS based on results from LDL-C GWAS of the discovery cohort followed by assessment of their performance in the replication cohort using linear regression models. The PGS panels tested during the optimization stage derived from different clumping LD r^2 and *P*-values thresholds are listed in supplemental Table S3. Figure 5 shows the performance values for these PGS panels tested in the replication cohort by adjusting for age, gender, cholesterol treatment and PCI-PC4 in the regression models. Correlation charts for these panels are presented in supplemental Figs. S1–S4. The clumping and thresholding method performed better compared to using the thresholding-only method (Fig. 5). The highest performing PGSs (PGS5_QBB_6SNP, PGS6_QBB_6SNP) were derived from the clumping and thresholding method (LD $r^2 = 0.2$) using a threshold of $P < 5.0 \times 10^{-10}$ and $P < 5.0 \times 10^{-12}$ (supplemental Table S3). The LDL-C prediction assessment of these best-performing 6-SNP panels

resulted in a higher adjusted R^2 value of 0.139 compared to an 83-SNP PGS derived without clumping using the same *P*-value threshold (adjusted R^2 value = 0.106; supplemental Table S4). Moreover, lower LD clumping r^2 threshold of 0.2 generally showed a superior performance compared to a higher r^2 value of 0.8. For example, thresholding at $P < 5.0 \times 10^{-12}$ and clumping at $r^2 = 0.8$ resulted in adjusted R^2 value of 0.134 compared with 0.139 at LD $r^2 = 0.2$ (Fig. 5). The top three performing PGS panels were selected for additional assessment and comparison with other previously derived PGS. PGS5_QBB_6SNP and PGS6_QBB_6SNP showed best performance (adj- $R^2 = 0.139$) and were found to contain the same SNPs. PGS5_QBB_6SNP and two additional PGS panels (PGS4_QBB_9SNP; Adj- $R^2 = 0.135$ and PGS14_QBB_14SNP; Adj- $R^2 = 0.134$) were selected for downstream comparisons (Table 2). The 3 best performing panels derived from the LDpred-2 were PGS25_QBB_269K_SNP (adj- $R^2 = 0.105$), PGS26_QBB_282K_SNP (adj- $R^2 = 0.107$) and PGS27_QBB_244K_SNP (adj- $R^2 = 0.103$).

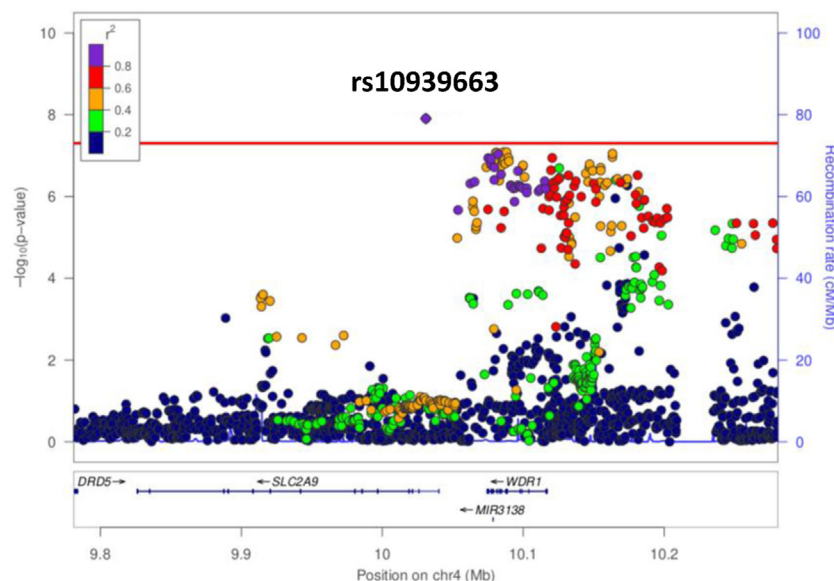


Fig. 4. Regional Association plot for the novel locus identified from the multi-ancestry meta-analysis for LDL-C. A novel locus identified from the multi-ancestry meta-analysis of the QBB cohort ($n = 13,701$) combined with the GLGC multi-ancestry cohort ($n = 1,654,960$). Variants are plotted using the GRCh38 build with x-axis representing the position in the chromosome and the y-axis representing the $-\log_{10}(P\text{-values})$. rs10939663 located in *SLC2A9*. The linkage disequilibrium (LD) reference from the local population was generated from the QBB data. The color of the plotted dots represents the LD r^2 value with the lead variants depicted as diamond-shaped markers. Recombination rates are also plotted with the corresponding value on the right y-axis. The red line represents $-\log_{10}$ of the genome-wide significance threshold ($P = 5.0 \times 10^{-8}$).

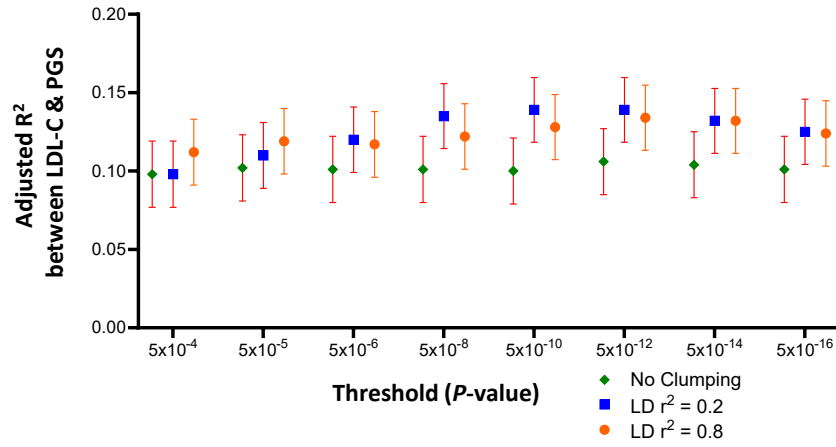


Fig. 5. Performance assessment of polygenic score (PGS) panels derived from discovery cohort ($n = 5,939$) and tested on the replication cohort ($n = 7,762$). Interleaved scatter plot shows the adjusted R^2 values from the linear regression models for raw LDL-C values with PGS derived using different P -value thresholds (x-axis) and clumping LD r^2 threshold values (0.2 or 0.8). The regression model included the PGS, age, gender, PCI-PC4, and cholesterol treatment as predictors. Data points represent mean adjusted R^2 with 95% Confidence Intervals (CI). Green diamond datapoints represent PGS derived based on thresholding only while PGS derived based on thresholding and clumping with $r^2 < 0.2$ are shown in blue and those with $r^2 < 0.8$ are shown in red.

Assessment of PGS panel performance

We compared the best-performing Qatari-derived LDL-C PGS panels against 11 previously developed LDL-C PGS panels from other ancestries (supplemental Table S5). These included the first LDL-C derived PGS from European ancestry that included 12 SNPs (EUR_12SNP; PGS000814 (15)), a revised version of this PGS that included 6 SNPs (EUR_6SNP; Refined PGS000814 (16)), a 36-SNP panel derived from European ancestry (EUR_36SNP; PGS000875 (34)). Additionally, we included the best performing LDL-C PGS panels derived from the largest GWAS of European

(EUR_1MNSNP; PGS000892, EUR_5427SNP: PGS000893), African (AFR_1MNSNP: PGS000886, AFR_29SNP; PGS000887,) and Multi-ancestry (MULTI_1MNSNP; PGS000888, MULTI_9009SNP; PGS000889) populations, along with two PGS from East Asian (EAS_66SNP; PGS000891) and South Asian (SAS_13SNP; PGS000897) populations from the same study (11). Notably, while all PGS panels included some common SNPs, the total number of SNPs/valid predictors in each PGS panel varied greatly. For instance, 5 of the 6 SNPs from the PGS5_QBB_6SNP occur in common genes as in the EUR_6SNP. The additional

TABLE 2. QBB-derived Polygenic score panels for LDL-C

PGS Panel	SNP	Closest Gene	A1	A2	Beta	A2 Freq in QBB Discovery	A2 Freq in IKGP (EUR)
PGS4_QBB_9SNP	rs35239705	<i>TDRD15</i>	A	G	0.125	0.76	0.78
	rs12609356	<i>SMARCA4</i>	C	T	0.109	0.63	0.64
PGS5_QBB_6SNP ^a	rs2738464 ^c	<i>LDLR</i>	G	C	0.143	0.82	0.89
	rs12740374	<i>CELSR2</i>	T	G	0.236	0.83	0.79
PGS14_QBB_14SNP	rs1800481 ^c	<i>APOB</i>	A	G	0.208	0.84	0.81
	rs144826254	<i>SMARCA4/LDLR</i>	G	T	0.239	0.83	0.88
	rs111234557	<i>MAU2</i>	G	C	0.264	0.93	0.89
	rs429358 ^b	<i>APOE</i>	C	T	-0.279	0.92	0.84
	rs7412 ^{b,c}	<i>APOE</i>	T	C	0.615	0.97	0.94
	rs4970834	<i>CELSR2</i>	T	C	0.218	0.86	0.81
	rs611917	<i>CELSR2</i>	G	A	0.201	0.77	0.69
	rs34722314	<i>APOB</i>	A	T	0.204	0.87	0.83
	rs562338	<i>APOB</i>	A	G	0.183	0.79	0.80
	rs9305019	<i>LDLR</i>	T	C	0.192	0.81	0.83
	rs10412048	<i>SMARCA4/LDLR</i>	G	A	0.201	0.80	0.89
	rs61679753	<i>TOMM40</i>	A	T	0.576	0.98	0.97
	rs141622900	<i>APOC1</i>	A	G	0.328	0.94	0.95

IKGP, 1000 Genomes Project (version 30X where available); A1, other allele; A2 effect allele; BETA, beta effect size; EUR, European population; Freq, Allele frequency; QBB, Qatar Biobank; SNP, Single Nucleotide Polymorphism.

^aSame as PGS6_QBB_6SNP.

^bVariants previously reported in the EUR_12SNP, EUR_6SNP panels.

^cVariants reported in the SAS_13SNP.

variant from the PGS5_QBB_6SNP occurs in *MAU2* (rs111234557), which is not represented in the EUR_6SNP nor the EUR_12SNP panels.

To assess the performance of LDL-C-derived PGS panels, we used linear regression models for LDL-C values and included the PGS, age, gender, cholesterol treatment and PCSk-4 as predictors, and as described in Graham *et al.* (11). Results showed that the best-performing panels were the multiethnic panels (MULTI_IMNSNP; adj- R^2 = 0.161 and MULTI_9009SNP panel; adj- R^2 = 0.160), followed by the largest European panels (EUR_IMNSNP; adj- R^2 = 0.156 and EUR_5427SNP; adj- R^2 = 0.141), and the Qatari-derived PGS5_QBB_6SNP (adj- R^2 = 0.139), PGS4_QBB_9SNP (adj- R^2 = 0.135) and PGS14_QBB_14SNP (adj- R^2 = 0.134) panels (Fig. 6A, supplemental Table S6, supplemental Figs. S4, and S5). The 3 PGS panels derived using the LDpred-2 method showed lower performance compared to PGS5_QBB_6SNP. Moreover, the SAS_13SNP (adj- R^2 = 0.132), AFR_295SNP (adj- R^2 = 0.126), and EAS_66SNP (adj- R^2 = 0.122) PGS performed better than the remaining 3 European-based SNP panels; EUR_6SNP, EUR_12SNP and EUR_36SNP. Interestingly, the top-performing PGSs showed similar performance when applied to the QBB cohort compared to previously derived PGS applied on their respective ancestries in the associated studies (Fig. 6B).

DISCUSSION

The present study represents the largest population-based GWAS to investigate the genomic determinants of LDL-C variation in Middle Eastern and Arab populations. Our GWAS replicated 8,043 variants located in 168 loci in the QBB dataset showing an overall consistent direction of effect indicating a significant trend for higher effect sizes in the QBB cohort. The observed variations in effect size estimates of individual variants based on pairwise comparisons between ancestries could indicate diverse patterns of linkage disequilibrium (LD); with the underlying causal variant, or a potential interaction with an environmental risk factor that exhibits varying prevalence across different ancestral backgrounds and/or geographical regions.

The multi-ancestry meta-analysis with the largest GWAS of LDL-C disclosed a novel locus in a gene that has not been previously associated with LDL-C concentration in blood. Our findings reflect the importance of the inclusion of Middle Eastern ancestral populations in GWAS since this novel locus exclusively reached genome-wide significance following the multi-ancestry meta-analysis. The identified LDL-C-associated novel locus was located in the intronic region of *SLC2A9*. This variant was initially below the threshold for genome-wide significance in the 7 cohorts (i.e., QBB discovery, QBB replication, and 5 Graham *et al.* single ancestry cohorts) but when combined,

it showed genome-wide significant association with LDL-C after genomic control correction. This gene has not been previously associated with LDL-C levels in the GWAS catalog (10), Phenoscanner (35) or the Common Metabolic Disease (CMD) Knowledge Portal (www.cmdkp.org).

The *Solute Carrier Family 2 Member 9* (*SLC2A9*) gene encodes a member of the SLC2A facilitative glucose transporter family (GLUT9), which is involved in the transportation of glucose and uric acid (36). Variants of *SLC2A9* are strongly associated with serum uric acid levels and gout (37, 38) in multiple ancestry individuals (39, 40). Importantly, while previous GWAS have predominantly associated variants in *SLC2A9* with metabolite levels, hypouricemia is also associated with serum lipid profiles (41, 42). Our findings present a variant in *SLC2A9* as a novel LDL-C-associated gene and its involvement in molecular pathways involved in cardiovascular disease progression and outcomes may be explored further. Combined, while our identified novel LDL-C-associated gene has been previously implicated with different pathological conditions, the impact of its genetic deviation on the gene regulatory network or molecular basis of LDL-C regulation, and associated risk of CVD development & progression warrant further investigations.

LDL-C levels can serve as early risk predictors of CVD and associated complications. PGS for LDL-C have been extensively explored in other populations to assess the genomic risk of hypercholesteremia and associated CVD risk, with varying performance and limited clinical utilization (43).

We derived Qatari-specific PGS panels for LDL-C prediction. While the genes included in our QBB PGS panels have also been previously associated with LDL-C levels, the specific set of variants has not been previously identified as the optimal representative panel to accurately predict LDL-C levels in other ancestries. For instance, in the PGS5_QBB_6SNP panel, 4 out of the 6 variants were specific to the Qatari population compared to the EUR_6/12SNP panel demonstrating the variation in genetic architecture between the two populations. Additionally, our identified variants occurred in genes similar to the EUR_6/12SNP panels (*LDLR*, *CELSR2*, *APOB*, and *APOE*) (15, 16). However, one variant in *MAU2* was only present in the Qatari PGS panel. Similarly, SAS_13SNP shared 6 of the 9 genes used in the QBB PGS panels, but only 3 SNPs; rs2738464, rs1800481, and rs7412 were shared between these panels, indicating genetic variance among populations.

In concordance with the Graham *et al.*, study, which reported that the MULTI_9009SNP PGS panel showed equivalent or better performance across most ancestry groups compared to the ancestry-specific PGS (11), the multi-ancestry derived panels also showed the best performance for LDL-C-prediction in our Qatari cohort. Polygenic scores derived from multiethnic

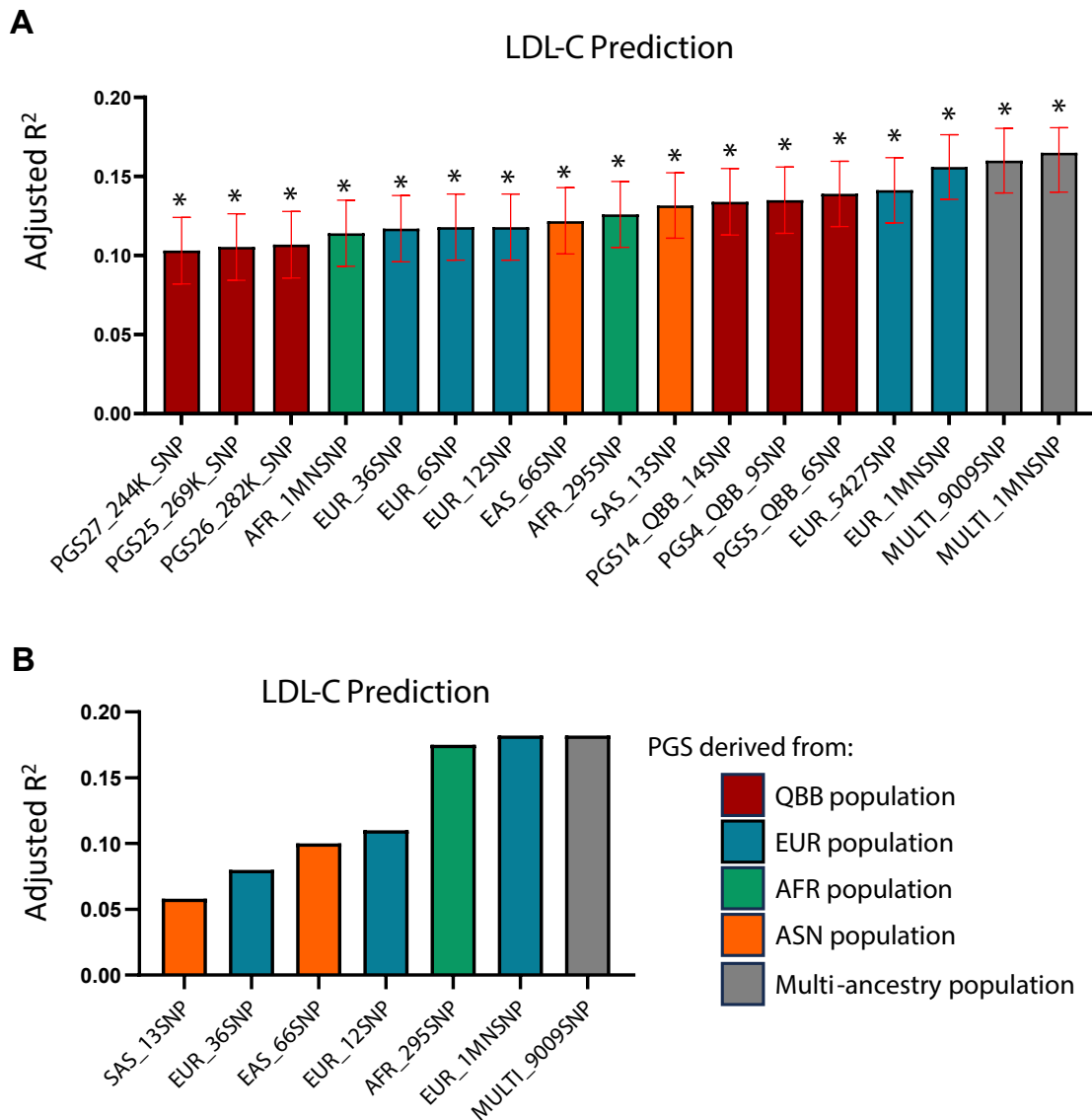


Fig. 6. Performance metrics for the polygenic score panels. A: Bar plot shows the adjusted R^2 values from the performance of the 17 different PGS panels tested on the QBB replication cohort ($n = 7,762$). Performance was assessed using linear regression for raw LDL-C values using PGS, age, gender, PCSL-4, and cholesterol treatment as predictors in the model. B: Bar plot shows the adjusted R^2 from the LDL-C linear regression models for different PGS panels ($n = 7$): EUR_36SNP ($n = 4,787$) (34), EUR_12SNP ($n = 3,020$) (15) and EUR_1MNSNP ($n = 389,158$) (11) tested on European populations, AFR_295SNP tested on African ancestry ($n = 6,863$), MULTI_9009SNP tested on multi-ancestry populations ($n = 461,918$), EAS_66SNP tested on East Asian populations ($n = 1,441$) and SAS_13SNP tested on South Asian populations ($n = 6,814$) (11). Asterisks (*) represent statistical significance for PGS from regression analysis ($P < 0.05$). QBB: Qatar Biobank, EUR: European, AFR: African, ASN: Asian, SAS: South Asian and EAS: East Asian populations.

ancestries can capture a wider range of genetic variations related to LDL-C and provide more accurate predictions for LDL-C traits. However, the performance metrics of the 3 Qatari-specific PGS were only slightly lower (0.022–0.027 lower adj- R^2) than the best performing multi-ancestry PGS. The SAS_13SNP PGS showed lower performance compared to the top 3 best performing QBB-derived panels. The EAS_66SNP also performed lower than the top three QBB panels, SAS_13SNP and AFR_295SNP, but better than the smaller European panels (EUR12SNP, EUR6SNP, EUR36SNP), larger African panel (AFR_1MNSNP) and

the QBB LDpred2-derived panels. Importantly, our best-performing QBB PGS panel comprised considerably fewer variants than the PGS panels from other ethnicities (6 variants in the Qatari panel compared to >1 million or ~5,000–9,000 variants in the multi-ancestry or European PGS panels), indicating the potent association of these sets of variants in LDL-C prediction within the Qatari population. Moreover, fewer SNPs in our PGS panels also support more feasible potential clinical utilization. Another study from the Middle East performed by the Tehran Cardiometabolic Genetic Study (TCGS) group (44), assessed

the predictive accuracy of a selected set of SNPs on lipid profile traits. The authors found that PGS based on their own ancestry outperformed PGS derived from other ancestries. In agreement with our study, they also reported that PGS panels from their ancestry with fewer SNPs showed comparable performance to PGS panels with a larger number of SNPs.

Our PGS panels were derived from whole genome sequence data providing more comprehensive coverage of genetic variations compared to genotyping arrays and imputation methods used in most previous GWAS. Of note, the EUR_12SNP and EUR_36SNP PGS showed stronger performances in predicting LDL-C variance in our study compared to their original studies. However, the analysis models used in the EUR_12SNP and EUR_36SNP studies were different from the model parameters used in our study, hindering accurate comparisons of performance metrics between the studies. The LDpred-2 models from our dataset did not perform as strongly as our C+T models possibly because of our small cohort size or because they need to be further optimized for our dataset (e.g. using a larger set of variants rather than restricted to common variants with HapMap). The benefits of this method are more obvious when dealing with larger sample sizes (45).

Although the findings from this study provide new insights into the genetic architecture of LDL-C, our study has some limitations. While the present study is the largest GWAS of LDL-C from a Middle Eastern population, it is relatively modest compared to recent published GWAS studies of LDL-C from European ancestry. However, in agreement with the findings of Graham *et al.*, the inclusion of individuals from diverse ancestries, such as the underrepresented Middle Eastern population, in meta-analysis leads to enhanced discoveries of novel variants and genomic regions for LDL-C as we demonstrated herein. Further larger studies from Middle Eastern populations will be required to capture more genetic variations from this ancestry.

Our study utilized population-based GWAS using WGS to identify variants that are associated with LDL-C levels within the Qatari population. We also performed a multi-ancestry meta-analysis with the largest LDL-C meta-analysis previously reported by the GLGC consortium (11). We replicated 168 genetic loci between QBB and Graham *et al.*, datasets and their allele frequencies and beta-effect values were broadly concordant. Importantly, we identified 1 novel LDL-C-associated locus in the multi-ancestry meta-analysis, indicating the importance of increasing ancestral diversity in GWAS. Additional investigations to decipher the molecular and pathophysiological roles of our identified variants in affecting LDL-C concentration in blood would facilitate the identification of therapeutic targets. The multi-ancestry-derived LDL-C PGS

(MULTI_IMNSNP) performed best at predicting LDL-C in our cohort. In addition, we developed 3 Qatari-based LDL-C PGS panels that performed reasonably well in predicting LDL-C levels. LDL-C level prediction has limited clinical utility since LDL-C levels are routinely tested in clinical practice. However, the identification of genetic variations and their effects on molecular pathways leading to variation in LDL-C levels has the potential for disclosing therapeutic targets. Moreover, the development of PGS to reflect a genetic predisposition to elevated LDL-C levels can aid in mitigating risk via the administration of medical or lifestyle interventions.

Data availability

The data analyzed in this study are subject to the following licenses/restrictions: the raw WGS data from the QBB are protected and are not available for deposition into public databases due to data privacy laws. Access to QBB/QGP phenotype and whole-genome sequence data can be obtained through an ISO-certified protocol, which involves submitting a project request at <https://www.qphi.org.qa/research/how-to-apply>, subject to approval by the Institutional Review Board of the QBB. ■

Supplemental data

This article contains [supplemental data](#) (11, 15, 16, 34).

Acknowledgment

We are grateful to the Qatar Biobank (QBB) and the Qatar Genome Program (QGP) for providing access to the genomic and phenotypic data. We are also thankful to all participants of this study.

Author contributions

S. M., M. S. H. A. H., and N. Q. resources; S. M. T. and U. I. U. writing—original draft; S. M. T. visualization; S. M. T., O. M. E. A., Y. A. A., and U. I. U. investigation; O. M. E. A. writing—review & editing; O. M. E. A. supervision; O. M. E. A. funding acquisition; O. M. E. A. and U. I. U. formal analysis; O. M. E. A. conceptualization; Y. A. A. methodology; Y. A. A. and U. I. U. data curation. N. Q., J. M. A. S., A. J., M. S., A. E., M. G., E. U., and K. K. resources.

Author ORCIDs

Umm-Kulthum Ismail Umlai  <https://orcid.org/0000-0002-4150-5027>

Salman M. Toor  <https://orcid.org/0000-0002-4683-3329>

Mohammed Gomaa  <https://orcid.org/0000-0002-6896-5760>

Nadeem Qureshi  <https://orcid.org/0000-0003-4909-0644>

Omar M.E. Albagha  <https://orcid.org/0000-0001-5916-5983>

Funding and additional information

This study was supported by a grant from the Qatar National Research Fund (QNRF) awarded to OMEA (PPM 03-0324-190038).

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations

CVD, Cardiovascular diseases; FH, familial hypercholesterolemia; GLGC, Global Lipids Genetics Consortium; GWAS, genome-wide association study; LD, linkage disequilibrium; LDL-C, low-density lipoprotein cholesterol; MAF, minor allele frequency; PGS, polygenic score; QBB, Qatar Biobank; WGS, Whole genome sequencing.

Manuscript received August 30, 2024, and in revised form December 12, 2024. Published, JLR Papers in Press, February 3, 2025, <https://doi.org/10.1016/j.jlr.2025.100752>

REFERENCES

- Ference, B. A., Ginsberg, H. N., Graham, I., Ray, K. K., Packard, C. J., Bruckert, E., *et al.* (2017) Low-density lipoproteins cause atherosclerotic cardiovascular disease. I. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* **38**, 2459–2472
- Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., *et al.* (2020) Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021
- Okraïneć, K., Banerjee, D. K., and Eisenberg, M. J. (2004) Coronary artery disease in the developing world. *Am. Heart J.* **148**, 7–15
- Almahmeed, W., Arnaout, M. S., Chettaoui, R., Ibrahim, M., Kurdi, M. I., Taher, M. A., *et al.* (2012) Coronary artery disease in Africa and the Middle East. *Ther. Clin. Risk Manag.* **8**, 65–72
- Baigent, C., Keech, A., Kearney, P. M., Blackwell, L., Buck, G., Pollicino, C., *et al.* (2005) Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet.* **366**, 1267–1278
- Cholesterol Treatment Trialists, C., Baigent, C., Blackwell, L., Emberson, J., Holland, L. E., Reith, C., *et al.* (2010) Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet.* **376**, 1670–1681
- Defesche, J. C., Gidding, S. S., Harada-Shiba, M., Hegele, R. A., Santos, R. D., and Wierzbicki, A. S. (2017) Familial hypercholesterolaemia. *Nat. Rev. Dis. Primers.* **3**, 17093
- Diboun, I., Al-Sarraj, Y., Toor, S. M., Mohammed, S., Qureshi, N., Al Hail, M. S. H., *et al.* (2022) The prevalence and genetic spectrum of familial hypercholesterolemia in Qatar based on whole genome sequencing of 14,000 subjects. *Front. Genet.* **13**, 927504
- Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283
- Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985
- Graham, S. E., Clarke, S. L., Wu, K.-H. H., Kanoni, S., Zajac, G. J. M., Ramdas, S., *et al.* (2021) The power of genetic diversity in genome-wide association studies of lipids. *Nature.* **600**, 675–679
- Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burt, N. P., Roos, C., *et al.* (2008) Polymorphisms associated with cholesterol and risk of cardiovascular events. *N. Engl. J. Med.* **358**, 1240–1249
- Isaacs, A., Willems, S. M., Bos, D., Dehghan, A., Hofman, A., Ikram, M. A., *et al.* (2013) Risk scores of common genetic variants for lipid levels influence atherosclerosis and incident coronary heart disease. *Arterioscler. Thromb. Vasc. Biol.* **33**, 2233–2239
- Trinder, M., Francis, G. A., and Brunham, L. R. (2020) Association of monogenic vs polygenic hypercholesterolemia with risk of atherosclerotic cardiovascular disease. *JAMA Cardiol.* **5**, 390–399
- Talmud, P. J., Shah, S., Whittall, R., Futema, M., Howard, P., Cooper, J. A., *et al.* (2013) Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet.* **381**, 1293–1301
- Futema, M., Shah, S., Cooper, J. A., Li, K., Whittall, R. A., Sharifi, M., *et al.* (2015) Refinement of variant selection for the LDL cholesterol genetic risk score in the diagnosis of the polygenic form of clinical familial hypercholesterolemia and replication in samples from 6 countries. *Clin. Chem.* **61**, 231–238
- Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., *et al.* (2021) The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* **53**, 420–425
- Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019) The missing diversity in human genetic studies. *Cell.* **177**, 26–31
- Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., *et al.* (2017) Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649
- Kamiza, A. B., Toure, S. M., Vujkovic, M., Machipisa, T., Soremekun, O. S., Kintu, C., *et al.* (2022) Transferability of genetic risk scores in African populations. *Nat. Med.* **28**, 1163–1166
- Kuchenbaecker, K., Telkar, N., Reiker, T., Walters, R. G., Lin, K., Eriksson, A., *et al.* (2019) The transferability of lipid loci across African, Asian and European cohorts. *Nat. Commun.* **10**, 4330
- Saad, M., El-Menyar, A., Kunji, K., Ullah, E., Al Suwaidi, J., and Kullo, I. J. (2022) Validation of polygenic risk scores for coronary heart disease in a middle eastern cohort using whole genome sequencing. *Circ. Genom. Precis. Med.* **15**, e003712
- Thareja, G., Al-Sarraj, Y., Belkadi, A., Almotawa, M., Qatar Genome Program Research, C., Suhre, K., *et al.* (2021) Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. *Nat. Commun.* **12**, 1250
- Al Thani, A., Fthenou, E., Paparrodopoulos, S., Al Marri, A., Shi, Z., Qafoud, F., *et al.* (2019) Qatar Biobank cohort study: study design and first results. *Am. J. Epidemiol.* **188**, 1420–1433
- Mbarek, H., Devadoss Gandhi, G., Selvaraj, S., Al-Muftah, W., Badji, R., Al-Sarraj, Y., *et al.* (2022) Qatar genome: insights on genomics from the Middle East. *Hum. Mutat.* **43**, 499–510
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498
- Slifer, S. H. (2018) PLINK: key functions for data analysis. *Curr. Protoc. Hum. Genet.* **97**, e59
- Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., *et al.* (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341
- Willer, C. J., Li, Y., and Abecasis, G. R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* **26**, 2190–2191
- Magi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N., *et al.* (2017) Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650
- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Glied, T. P., *et al.* (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* **26**, 2336–2337
- Choi, S. W., Mak, T. S., and O'Reilly, P. F. (2020) Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772
- Prive, F., Arbel, J., and Vilhjalmsón, B. J. (2021) LDpred2: better, faster. *Stronger. Bioinformatics.* **36**, 5424–5431
- Leal, L. G., Hoggart, C., Jarvelin, M.-R., Herzog, K.-H., Sternberg, M. J. E., and David, A. (2020) A polygenic biomarker to identify patients with severe hypercholesterolemia of polygenic origin. *Mol. Genet. Genomic Med.* **8**, e1248
- Staley, J. R., Blackshaw, J., Kamat, M. A., Ellis, S., Surendran, P., Sun, B. B., *et al.* (2016) PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics.* **32**, 3207–3209
- Le, M. T., Shafiu, M., Mu, W., and Johnson, R. J. (2008) SLC2A9-a fructose transporter identified as a novel uric acid transporter. *Nephrol. Dial. Transpl.* **23**, 2746–2749

37. Doring, A., Gieger, C., Mehta, D., Gohlke, H., Prokisch, H., Coassin, S., *et al.* (2008) SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat. Genet.* **40**, 430–436
38. Vitart, V., Rudan, I., Hayward, C., Gray, N. K., Floyd, J., Palmer, C. N., *et al.* (2008) SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat. Genet.* **40**, 437–442
39. Tin, A., Marten, J., Halperin Kuhns, V. L., Li, Y., Wuttke, M., Kirsten, H., *et al.* (2019) Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.* **51**, 1459–1474
40. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshihara, S., *et al.* (2021) A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424
41. Peng, T. C., Wang, C. C., Kao, T. W., Chan, J. Y., Yang, Y. H., Chang, Y. W., *et al.* (2015) Relationship between hyperuricemia and lipid profiles in US adults. *Biomed. Res. Int.* **2015**, 127596
42. Qi, J., Dai, X., Zhou, B., Su, Y., Xie, Z., and Liu, D. (2021) Association between lipid profiles and serum urate: a cross-sectional study in southwestern China. *Int. J. Endocrinol.* **2021**, 2741131
43. Cupido, A. J., Tromp, T. R., and Hovingh, G. K. (2021) The clinical applicability of polygenic risk scores for LDL-cholesterol: considerations, current evidence and future perspectives. *Curr. Opin. Lipidol.* **32**, 112–116
44. Akbarzadeh, M., Dehkordi, S. R., Roudbar, M. A., Sargolzaei, M., Guity, K., Sedaghati-Khayat, B., *et al.* (2021) GWAS findings improved genomic prediction accuracy of lipid profile traits: Tehran Cardiometabolic Genetic Study. *Sci. Rep.* **11**, 5780
45. Vilhjalmsen, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., *et al.* (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592