Supplemental Material

Supplemental Methods

Data sources and availability
The low-pass whole-genome sequencing (WGS) dataset (phase 1) and the genotyping dataset (phase 2) of the MHI Biobank have been previously described \(^1,^2\). Case status for coronary artery disease (CAD) prevalence was defined as having a myocardial infarction (MI) or coronary artery interventions (coronary artery bypass grafting (CABG) or percutaneous coronary intervention (PCI)) before the first visit. Controls were selected to be free of MI, PCI, CABG, transient ischemic attack or stroke, peripheral vascular disease, congestive heart failure, and angina. We used the same clinical definitions to identify incident and recurrent CAD events in the MHI Biobank: participants who had a first-ever CAD event between baseline and follow-up were considered incident cases and participants who had at least a second CAD event during the same period were considered recurrent CAD cases. Average follow-up time was 4.2 years for phase 1 and 3.6 years for phase 2. We used GenomeStrip (v.2.0) with default parameters on the MHI Biobank phase 1 WGS data to identify participants who carry the known French-Canadian founder \(LDLR\) deletion \(^3\). We excluded the 14 individuals who carry the \(LDLR\) delta > 15 kb deletion from all subsequent PRS analyses.

CARTaGENE (www.cartagene.qc.ca) is a population-based cohort of Quebec that includes individuals aged between 40 and 69 years \(^4\). A subset of this cohort totaling 5762 samples were genotyped on the Illumina Infinium Global Screening Array (GSA). We used PLINK (version 1.9, https://www.cog-genomics.org/plink/1.9/) to apply the following quality-control filters: we excluded samples and variants with >5% missingness, variants out of Hardy-Weinberg Equilibrium (p-value<1x10\(^{-6}\)), A/T and G/C variants, and variants with a minor allele frequency (MAF) <1%. Following these quality-control steps, we phased genotypes with ShapeIT v2.r790 and imputed missing genotypes on the Michigan Imputation Server (version 1.30.4) using the Haplotype Reference Consortium panel (Version r1.1 2016) \(^5,^6\). Case-control status for CAD in CARTaGENE was defined using the same criteria than for the MHI Biobank (see above). Principal components for all three cohorts were calculated in PLINK using the pca function.

Polygenic risk scores
The models for the two polygenic risk scores (PRS) (GPS\(_{CAD}\) and metaGRS\(_{CAD}\)) used in this study are available online (see Web links below) \(^7,^8\). Genetic risk scores for all models were generated with PLINK (version 1.9) and the --score function to calculate the sum of the product of the dosage of effect alleles per variant weighted by the CAD effect size \(^9\). We excluded variants with low imputation quality score (rsq <0.3). PRS were Z-score normalized and centered (mean=0, standard deviation=1) per dataset to facilitate interpretation of odds ratios. Detailed numbers for missing variants from all models can be found in Table S1.
Statistical analysis
We performed all statistical analysis in R (version 3.5.0)\(^{10}\). In the MHI Biobank datasets, we tested by logistic regression the association between CAD case-control status and PRS Z-scores correcting for age, sex, and the first four principal components. In the MHI Biobank phase 1 and 2 data, we also corrected for statin use when appropriate. For the CARTaGENE data, we used a similar logistic regression model, correcting for age, sex, the first four principal components, and recruitment center. We calculated odds ratios per standard deviation of the \( \text{GPS}_{\text{CAD}} \) or \( \text{metaGRS}_{\text{CAD}} \) PRS, and considered \( p \)-value < 0.05 as significant. P-values were not corrected for multiple testing. We calculated the area under the curve (AUC) using the \text{pROC} package in R\(^{11}\). Meta-analysis was performed with the R \text{metafor} package using the coefficients and standard errors (SE) from the individual regression models with a fixed effect model fitted with the “FE” method\(^ {12}\). We tested the association between LDL-cholesterol levels and the \( \text{LDLR} \) deletion in RVTest using the \text{score test} function\(^ {13}\). For these analyses, we increased the LDL-C levels of dyslipidemic participants by 30% to account for the effect of statins.

Supplemental References:

1. Low-Kam, C. \textit{et al.} Variants at the APOE /C1/C2/C4 Locus Modulate Cholesterol Efflux Capacity Independently of High-Density Lipoprotein Cholesterol. \textit{J Am Heart Assoc.} 2018;7:e009545.

2. Low-Kam, C. \textit{et al.} Whole-genome sequencing in French Canadians from Quebec. \textit{Hum Genet.} 2016;135:1213–1221.

3. Handsaker, R. E. \textit{et al.} Large multiallelic copy number variations in humans. \textit{Nat Genet.} 2015;47:296–303.

4. Awadalla, P. \textit{et al.} Cohort profile of the CARTaGENE study: Quebec’s population-based biobank for public health and personalized genomics. \textit{Int J Epidemiol.} 2013;42:1285–1299.

5. Das, S. \textit{et al.} Next-generation genotype imputation service and methods. \textit{Nat Genet.} 2016;48:1284–1287.

6. McCarthy, S. \textit{et al.} A reference panel of 64,976 haplotypes for genotype imputation. \textit{Nat Genet.} 2016;48:1279–1283.

7. Khera, A. V. \textit{et al.} Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. \textit{Nat Genet.} 2018;50:1219–1224.

8. Inouye, M. \textit{et al.} Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. \textit{J Am Coll Cardiol.} 2018;72:1883–1893.

9. Purcell, S. \textit{et al.} PLINK: a tool set for whole-genome association and population-based linkage analyses. \textit{Am J Hum Genet.} 2007;81:559–575.

10. Bates, D. \textit{et al.} R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2018.
11. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* 2011;12:77.

12. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2009;36:1–48.

13. Zhan X, et al. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinforma.* 2016;32:1423–1426.
**Supplemental Tables**

**Table S1.** Variants available from each cohort to calculate the two coronary artery disease polygenic risk scores. PRS = polygenic risk score used, n variants model = Total variants included in the model, variants scored = Percentage of variants from model that were present in data and used for scoring, variants missing = percentage of variants missing from scoring, rsq < 0.3 = number of variants that were excluded due to low imputation quality, other ALT = number of variants excluded due to different alternative allele than in reference model, missing = number of PRS model variants absent in dataset. N.A. = not applicable.

| Cohort            | PRS   | n variants model | variants scored (%) | variants missing (%) | rsq < 0.3 | other ALT | missing |
|-------------------|-------|------------------|---------------------|----------------------|------------|-----------|---------|
| MHI Biobank phase1 | GPS<sub>CAD</sub> | 6630150           | 94.01               | 5.99                 | N.A.       | 5394      | 391674  |
| MHI Biobank phase2 | GPS<sub>CAD</sub> | 6630150           | 93.04               | 6.96                 | 47323      | 50        | 414406  |
| CARTaGENE         | GPS<sub>CAD</sub> | 6630150           | 94.46               | 5.54                 | 1475       | 3203      | 363401  |
| MHI Biobank phase1 | metaGRS<sub>CAD</sub> | 1745179           | 97.55               | 2.45                 | NA         | 616       | 42124   |
| MHI Biobank phase2 | metaGRS<sub>CAD</sub> | 1745179           | 96.16               | 3.84                 | 33351      | 85        | 33595   |
| CARTaGENE         | metaGRS<sub>CAD</sub> | 1745179           | 99.91               | 0.09                 | 1475       | 50        | 0       |
Table S2. Association of polygenic risk scores with incident CAD for the MHI Biobank cohorts phase 1 and 2. N.A. = not applicable.

| Model        | Cohort                         | Phenotype     | Cases | Controls | p-value       | Odds ratio   | OR (95% CI) | AUC | AUC (95% CI) |
|--------------|--------------------------------|---------------|-------|----------|---------------|--------------|-------------|-----|-------------|
| **GPS\textsubscript{CAD}** | MHI Biobank phase 1           | CAD incidence | 257   | 636      | 4.76E-03      | 1.25         | 1.07-1.45   | 0.60| 0.51-0.62   |
|              | MHI Biobank phase 2           | CAD incidence | 145   | 609      | 0.57          | 0.95         | 0.79-1.40   | 0.57| 0.51-0.62   |
| Meta-analysis| CAD incidence                 | 402           | 1245  | 0.071    | 1.11          | 0.99-1.25    | N.A.        | N.A.|
| **metaGRS\textsubscript{CAD}** | MHI Biobank phase 1           | CAD incidence | 257   | 636      | 0.05          | 1.17         | 1.00-1.37   | 0.59| 0.55-0.63   |
|              | MHI Biobank phase 2           | CAD incidence | 145   | 609      | 0.22          | 1.12         | 0.93-1.35   | 0.57| 0.52-0.62   |
| Meta-analysis| CAD incidence                 | 402           | 1245  | 0.022    | 1.15          | 1.02-1.29    | N.A.        | N.A.|

N.A. = not applicable.