Title
Genetic architecture of cardiometabolic risks in people living with HIV.

Permalink
https://escholarship.org/uc/item/0sv5n1vg

Journal
BMC medicine, 18(1)

ISSN
1741-7015

Authors
Cheng, Haoxiang
Sewda, Anshuman
Marquez-Luna, Carla
et al.

Publication Date
2020-10-28

DOI
10.1186/s12916-020-01762-z

Peer reviewed
Genetic architecture of cardiometabolic risks in people living with HIV

Haoxiang Chang†, Anshuman Sewda†,2, Carla Marquez-Luna3, Sierra R. White1, Bridget M. Whitney6, Jessica Williams-Nguyen9, Robin M. Nance1,5, Won Jun Lee1, Mari M. Kitahata5,6, Michael S. Saag7, Amanda Willig7, Joseph J. Eron8, W. Christopher Mathews9, Peter W. Hunt10, Richard D. Moore11,12, Allison Webel13, Kenneth H. Mayer14, Joseph A. Delaney4, Paul K. Crane5, Heidi M. Crane5,6, Ke Hao1† and Inga Peter1*†

Abstract

Background: Advances in antiretroviral therapies have greatly improved the survival of people living with human immunodeficiency virus (HIV) infection (PLWH); yet, PLWH have a higher risk of cardiovascular disease than those without HIV. While numerous genetic loci have been linked to cardiometabolic risk in the general population, genetic predictors of the excessive risk in PLWH are largely unknown.

Methods: We screened for common and HIV-specific genetic variants associated with variation in lipid levels in 6284 PLWH (3095 European Americans [EA] and 3189 African Americans [AA]), from the Centers for AIDS Research Network of Integrated Clinical Systems cohort. Genetic hits found exclusively in the PLWH cohort were tested for association with other traits. We then assessed the predictive value of a series of polygenic risk scores (PRS) recapitulating the genetic burden for lipid levels, type 2 diabetes (T2D), and myocardial infarction (MI) in EA and AA PLWH.

Results: We confirmed the impact of previously reported lipid-related susceptibility loci in PLWH. Furthermore, we identified PLWH-specific variants in genes involved in immune cell regulation and previously linked to HIV control, body composition, smoking, and alcohol consumption. Moreover, PLWH at the top of European-based PRS for T2D distribution demonstrated a > 2-fold increased risk of T2D compared to the remaining 95% in EA PLWH but to a much lesser degree in AA. Importantly, while PRS for MI was not predictive of MI risk in AA PLWH, multiethnic PRS significantly improved risk stratification for T2D and MI.

Conclusions: Our findings suggest that genetic loci involved in the regulation of the immune system and predisposition to risky behaviors contribute to dyslipidemia in the presence of HIV infection. Moreover, we demonstrate the utility of the European-based and multiethnic PRS for stratification of PLWH at a high risk of cardiometabolic diseases who may benefit from preventive therapies.

Keywords: HIV, Polygenic risk score, Lipoprotein, Triglyceride, Type 2 diabetes, Myocardial infarction, Genome-wide association study
Background
The number of people living with human immunodeficiency virus (HIV) infection (PLWH) worldwide has increased by 34.6% (from 27.4 million to 36.9 million) between 2000 and 2018, while acquired immune deficiency syndrome (AIDS)-related deaths have declined from 1.5 million to 940,000 annually [1]. These advances can be primarily attributed to therapeutic advances in antiretroviral therapy (ART) and improved access to ART, allowing PLWH to live longer. However, accumulating evidence suggests that PLWH are at a higher risk of cardiovascular diseases (CVD) and have increased CVD-related mortality rates than those without HIV [2–6]. The possible causes of increased CVD risk among PLWH include inflammation and immune activation in response to HIV infection and viremia, adverse effects of ART, and lifestyle risk factors (e.g., smoking, alcohol, and illicit drug use). However, these factors do not fully account for the increased risk of CVD in PLWH [7, 8].

Genetic variants have been identified as significant predictors of traditional CVD risk factors including cardiometabolic traits and diseases, such as dyslipidemia and lipid levels (low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), and triglycerides) [9, 10], obesity [11, 12], type 2 diabetes mellitus (T2D) [13], and myocardial infarction (MI) [14] in the general population. CVD and related disorders have been demonstrated to have polygenic modes of inheritance, meaning that common genetic variants with small effect sizes located in multiple genes contribute to variability in disease or trait risk [15, 16]. Polygenic risk scores (PRS) have been proposed to assess the cumulative burden of multiple common susceptibility loci [17, 18]. A recent study found that 8% of the population possesses a genetic predisposition that confers a more than three-fold increased risk for coronary artery disease (CAD), with the highest PRS percentiles identifying 20 times more people than found by familial hypercholesterolemia mutations at a comparable or higher risk [19–21]. Moreover, in randomized clinical trials, people with the highest burden of genetic risk demonstrated the most substantial clinical benefit from primary prevention (statin therapy) resulting in a roughly three-fold decrease in the number needed to treat to prevent one CAD event [22].

Despite the growing literature proposing the clinical value of PRS in the general population [23], only a few reports with limited sample sizes have demonstrated the contribution of genetic variation to cardiometabolic risk in PLWH [24–26]; even fewer have examined the utility of PRS in PLWH [27]. Therefore, this study aimed to identify genetic predictors of cardiometabolic traits in PLWH and systematically assess the performance of PRS derived using results from previously published well-powered genome-wide association studies (GWAS) of T2D [28], CAD [29, 30], lipids (LDL, HDL, and triglyceride levels) [31], and body mass index (BMI) [32], and genomic data from the largest ethnically diverse PLWH cohort to date with genetic information. Given the emerging interest in applying PRS to improve clinical decision making [33], this study may help shed light on the genetic predictors of cardiometabolic risk in the presence of HIV infection and improve risk stratification to identify individuals at a high risk of CVD.

Methods
Study participants
The Centers for AIDS Research Network of Integrated Clinical Systems (CNICS) cohort includes a multiethnic population of ~36,000 PLWH (age 18 years and older) who have received routine clinical care at one of eight sites in the USA [34]. CNICS has an ongoing genetics project in which adult PLWH across racial/ethnic backgrounds from all sites, who provided informed consent and contributed specimens to the CNICS biospecimen repository, are being genotyped. Study participants were included if their genetic data were available at the time of these analyses.

Measurement of cardiometabolic phenotypes
The CNICS data repository integrates comprehensive clinical data from sites from outpatient and inpatient encounters, including information on demographic characteristics, clinical and laboratory data, medications, and historical clinical information. Lipid levels in CNICS include HDL, LDL, and triglyceride values measured as part of routine care and, therefore, may or may not have been obtained in the fasting state. LDL was either measured directly or calculated using the Friedewald equation [35]. BMI was calculated from heights and weights as a continuous variable (kg/m²). PLWH were categorized as ART-naïve or experienced. Among participants, the initial CNICS visit dates ranged from 1995 to 2015. Between the initial and the last CNICS visits, the average follow-up period was 10.3 years (median, 9.9 years; range, 0–23 years). Most included PLWH had multiple recorded values for each lipid drawn as part of care, we used mean values. We excluded individuals who were taking lipid-lowering drugs (e.g., HMG Co-A reductase inhibitors or statins) at baseline.

T2D diagnosis in CNICS is based on the following criteria: (1) hemoglobin A1c ≥ 6.5; (2) use of a diabetes-specific medication such as insulin; or (3) use of a diabetes-related medication, which is frequently, but not exclusively, used to treat diabetes (e.g., biguanides) in the setting of also having a diabetes diagnosis [36]. We have found high sensitivity (99%) and specificity (97%) for this definition [36].
CNICS uses an established state-of-the-art approach to adjudicate [37, 38] and classify MIs based on the Universal Definition of MIs [39, 40]. Potential MIs in the centralized CNICS data repository were identified using a comprehensive set of MI diagnostic and procedure codes and elevated cardiac biomarker values to optimize the ascertainment sensitivity as previously described [37, 38]. De-identified packets were prepared that contained provider notes, electrocardiograms, laboratory reports, and results from imaging and procedures, such as cardiac catheterization. Two physicians with expertise in adjudicating cardiac events performed a centralized review of the patient data, followed by inputs from a third physician for resolving discrepancies. We included type 1 MIs, those due to atheroembolic disease, and excluded type 2 MIs due to a mismatch in the oxygen supply and demand, usually observed in the setting of sepsis or cocaine or other illicit drug-induced vasospasm [37].

Genotyping and imputation
DNA was isolated from peripheral blood mononuclear cells or buffy coats of PLWH obtained from the CNICS biorepository using the FlexiGene DNA kit (Qiagen, #51206). DNA samples were then normalized and genotyped using Illumina’s high-density custom Multiethnic Global Array (MEGA) series BeadChips. Genotyped variant calling was performed using GenomeStudio® Genotyping Module v2.0 software (Illumina®, San Diego, California, USA) and zCall [41]. PLINK v.1.9 was used to exclude single nucleotide polymorphisms (SNPs) with call rates < 95%, minor allele frequency < 1%, and deviation from Hardy-Weinberg equilibrium (p value < 1E-5), as well as samples with call rates < 90%, sex discrepancies between genotype data and self-report, and pairwise identity-by-descent (pi-hat > 0.9) [42].

We inferred ethnicity on genotype data using GRAPPop software [43], and, after excluding the human leukocyte antigen encoding region, performed principal components analysis (PCA) on the African American (AA) and European American (EA) samples separately using EIGENSOFT [44]. The estimated principal components (PCs) were included in the regression models while performing genome-wide association analysis in each ancestry group. Genotype data from each ancestry group was imputed separately using the cloud-based Michigan Imputation Server [45] and Trans-Omics for Precision Medicine, or TOPMed data, as the reference panel (https://www.nhlbiwgs.org/). For further analysis, we only kept variants that were imputed with high quality (imputation quality score, r² > 0.3) and passed the standard quality control procedures. The genotyped and imputed SNP counts are listed in Additional file 1: Table S1.

Genome-wide association analysis
Genome-wide association tests were conducted on each SNP using either linear or logistic regression method on imputed dosage data sets, using in-house code written in R (version 3.5.3). The tests were performed separately in European and African ancestry sub-cohorts, and then pooled using random-effects meta-analysis, implemented in the “meta” R package [46]. In addition to the first ten PCs, analyses were adjusted for site, age, sex at birth, and presence or absence of ART. A study reported that genetic associations with lipid traits differed by sex [47]; therefore, we repeated these analyses in male and female sub-cohorts separately. The results were visualized through multi-phenotype and single-phenotype mirrored Manhattan plots. HIV-specific genetic variants were defined as loci that were significant at p < 0.01 in GWASHV and had p ≥ 0.05 in the well-powered GWASGEN and the 99% confidence intervals (CI) for the beta coefficients in GWASHV and GWASGEN did not overlap. Similar approach was used to detect ancestry-specific or sex-specific lipid-related variants.

Gene set enrichment analysis
Enrichr was used to perform gene set enrichment analyses using the genes containing HIV-specific variants. Enrichr database is an integrative web-based application, currently containing 335,434 annotated gene sets from 166 gene set libraries [48, 49]. UK Biobank consists of a large prospective cohort of more than 500,000 middle-aged participants with detailed information on a wide range of complex diseases, lifestyle risk factors, medical history, and physical measurements [50]. The health outcomes were adjudicated by experts for a range of disease areas. The genetic data and statistical analyses were synchronized across multiple phenotypes. We looked for enrichment in the UK Biobank GWAS version 1 (https://www.ukbiobank.ac.uk/tag/gwas/) gene set library which contains 857 terms covering 14,148 genes (122 genes per term). Adjusted p values calculated using the false discovery rate (FDR) for correction for multiple hypotheses testing [51] were reported for each term. An adjusted p < 0.05 was considered statistically significant.

Expression quantitative trait loci (eQTL) analysis
To assess the functional relevance of the newly observed associations, we tested whether HIV-specific loci are enriched among variants shown to regulate gene expression (eQTLs). We acquired eQTL data in primary CD14+ human monocytes from 432 European volunteers at baseline and after exposure to the inflammatory proxies interferon-γ (IFN-γ) or differing durations (2 h or 24 h) of lipopolysaccharide (LPS), which was profiled using the Illumina Human OmniExpress BeadChips genotyping array [52]. SNPs that were significantly
associated with each trait at $p < 10^{-6}$ in GWAS\textsubscript{GEN} of lipid profiles were excluded [31]. Furthermore, linkage disequilibrium (LD)-based pruning was performed using a threshold of $r^2 > 0.2$. After variant-filtering, we used chi-squared tests to compare the proportion of the eQTL SNPs (eSNPs) that were associated with gene expression levels at 10% FDR, among the HIV-specific loci to the remaining non-significant SNPs.

**Polygenic risk score analysis**

**Traditional PRS**

The PRS, representing estimated genetic determinants for five traits (HDL, LDL, triglycerides, T2D, and type 1 MI), were computed following the thresholding-pruning procedure [53]. We computed PRS for EA sub-cohort of PLWH (PLWH\textsubscript{EA}) and AA sub-cohort of PLWH (PLWH\textsubscript{AA}) separately using linear combinations of the imputed genotype dosages [54], and regression coefficients from the respective summary association statistics retrieved from previously published GWAS conducted in the general population largely of European ancestry: Global Lipids Genetics Consortium (GLGC) [31]; Genetic Investigation of ANthropometric Traits (GIANT) consortium [32]; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium [28]; Coronary ARtery Disease Genome wide Replication and Meta-analysis (DIAGRAM) consortium [30] (PRS\textsubscript{GEN}, Additional file 1: Table S2). For each disease/trait, we calculated eight sets of PRS from previously published GWAS conducted in the general population of European ancestry: Global Lipids Genetics Consortium (GLGC) [31]; Genetic Investigation of ANthropometric Traits (GIANT) consortium [32]; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium [28]; Coronary ARtery Disease Genome wide Replication and Meta-analysis (DIAGRAM) consortium [30].

**Multiethnic PRS**

To derive PRS that would perform well for both PLWH\textsubscript{EA} and PLWH\textsubscript{AA}, we considered GWAS summary statistics from two training sources: (1) the GWAS conducted in the general population of European ancestry (PRS\textsubscript{EA}) and (2) the GWAS conducted in PLWH\textsubscript{AA} (PRS\textsubscript{AA}), using ten-fold cross-validation. Additionally, we derived multiethnic PRS (Additional file 1: Table S3) that combined the two training sources using a recently published method [56]. Briefly, the multiethnic PRS is defined as the linear combination of the two PRSs with mixing weights $\alpha_1$ and $\alpha_2$. That is,

$$\text{PRS}_{\text{EA+AA}} = \alpha_1 \text{PRS}_{\text{EA}} + \alpha_2 \text{PRS}_{\text{AA}}$$

We estimated mixing weights $\alpha_1$ and $\alpha_2$ using validation data by fitting a linear regression model and computed adjusted $R^2$ to account for the additional degree of freedom. We employed a ten-fold cross-validation, using 90% of the cohort to estimate GWAS regression coefficients and the remaining 10% of the cohort to validate predictions (using the adjusted-$R^2$ metric with best-fit mixture weights, $\hat{\alpha}_1$ and $\hat{\alpha}_2$) and reported an average adjusted $R^2$ across the ten cross-validations. For each fold, we computed regression coefficients using linear regression for quantitative traits while adjusting for 10 PCs, sex, age, age$^2$, presence or absence of ART, and site, where the PCs were estimated using only PLWH\textsubscript{AA}. For T2D and MI diagnoses that had low prevalence in our cohort, we used stratified ten-fold cross-validation, where each cross-validation had the same case-control ratio. For lipid traits, for each $p$ value threshold, we calculated the $R^2$ statistic derived from a fixed-effects meta-analysis of marginal associations between PRS\textsubscript{EA+AA} and the trait of interest.

Lastly, we estimated the prevalence of T2D and MI for PLWH with the highest European-based and multiethnic PRS. We applied multiple testing correction to account for the number of thresholds and PRS tested using FDR [51]. An adjusted $p < 0.05$ was considered statistically significant. The number of SNPs used to calculate various multiethnic PRS is reported in Additional file 1: Table S3.

**Results**

The final cohort consisted of 6284 PLWH with 3095 PLWH\textsubscript{EA} and 3189 PLWH\textsubscript{AA}; both sub-cohorts were predominantly male (89% and 69%, respectively), which is consistent with the HIV epidemic in the USA (Table 1). PLWH\textsubscript{AA} had a higher prevalence of T2D ($p < 0.0001$, Table 1), but lower mean LDL ($p < 0.0001$) and triglyceride ($p < 0.0001$) levels and higher mean HDL levels ($p < 0.0001$) than PLWH\textsubscript{EA} (Table 2).

Figure 1 summarizes GWAS results for HDL, LDL, and triglycerides in PLWH\textsubscript{EA} alongside previously reported findings in populations of European ancestry [31]. We confirmed strong associations exceeding genome-wide statistical significance of variation in APOE (apolipoprotein E), CETP (Cholesteryl Ester Transfer Protein) with HDL levels; APOE and APOC1 (apolipoprotein C1) with LDL levels, and APOA5 (apolipoprotein A5), BUD13 (BUD13 Homolog), and TRIB1 (Tribbles Pseudokinase 1) with triglyceride levels in PLWH\textsubscript{EA}.
Additional associations at $p < 1 \times 10^{-5}$ in both HIV and no-HIV cohorts were detected in other previously reported lipid-related genes, including $LIPC$ (Lipase C) and $AQP9$ (Aquaporin 9) for HDL; $NECTIN2$ (Nectin Cell Adhesion Molecule 2), $CELSR2$ (Cadherin EGF LAG Seven-Pass G-Type Receptor 2), $PSRC1$ (Proline And Serine Rich Coiled-Coil 1), $APOC4$-$APOC2$ (apolipoprotein C4, C2), and $TOMM40$ (Translocase Of Outer Mitochondrial Membrane 40) for LDL; and $LPL$ (Lipoprotein Lipase), $ZPR1$ (Zinc Finger Protein 259), and $SLC18A1$ for triglycerides (Fig. 1, top panel; Additional file 2: Table S4). Furthermore, we identified variants that were significant in GWAS HIV but not in GWAS GEN, despite

### Table 1 Baseline demographic and clinical characteristics of the study cohort

| Variable | PLWH_{EA} | PLWH_{AA} | Total | $p$ value$^c$ |
|----------|------------|------------|-------|--------------|
| Age      | 53.36 ± 9.70 | 53.18 ± 10.70 | 53.27 ± 10.22 | 0.49 |
| Gender   | < 0.0001 |
|          | Male 2763 (89.3) | 2199 (69.0) | 4962 (79.0) | |
|          | Female 332 (10.7) | 990 (31.0) | 1322 (21.0) | |
| Site     | < 0.0001 |
|          | University of Alabama 744 (24.0) | 882 (27.7) | 1626 (25.9) | |
|          | Johns Hopkins 135 (4.4) | 845 (26.5) | 980 (15.6) | |
|          | University of Washington 623 (20.1) | 261 (8.2) | 884 (14.1) | |
|          | University of California San Diego 640 (20.7) | 187 (5.9) | 827 (13.2) | |
|          | Case Western Reserve University 314 (10.1) | 494 (15.5) | 808 (12.9) | |
|          | University of North Carolina 161 (5.2) | 368 (11.5) | 529 (8.4) | |
|          | Fenway 309 (10.0) | 45 (1.4) | 354 (5.6) | |
|          | University of California San Francisco 169 (5.5) | 107 (3.4) | 276 (4.4) | |
| Type 2 diabetes$^d$ | 388 (12.5) | 676 (21.2) | 1064 (16.9) | < 0.0001 |
| Myocardial infarction$^e$ | 53 (1.7) | 64 (2.0) | 117 (1.9) | < 0.39 |
| CD4 counts$^c$ | 399 ± 283.5 | 331 ± 277.3 | 364 ± 282.4 | < 0.0001 |
| Presence of antiretroviral therapy | 2841 (91.8) | 2823 (88.5) | 5664 (90.1) | < 0.0001 |

$^a$PLWH_{EA}, European American sub-cohort of people living with HIV. $^b$PLWH_{AA}, African American sub-cohort of people living with HIV. $^c$The $p$ values were calculated using a t-test. $^d$During study follow-up. $^e$At baseline

(Fig. 1, top panel; Additional file 2: Table S4). Additional associations at $p < 1 \times 10^{-5}$ in both HIV and no-HIV cohorts were detected in other previously reported lipid-related genes, including $LIPC$ (Lipase C) and $AQP9$ (Aquaporin 9) for HDL; $NECTIN2$ (Nectin Cell Adhesion Molecule 2), $CELSR2$ (Cadherin EGF LAG Seven-Pass G-Type Receptor 2), $PSRC1$ (Proline And Serine

### Table 2 Mean (standard deviation) and mean comparison $p$ values for lipid values stratified by European American vs. African American race in the study cohort

| Trait | Subgroup | PLWH_{EA} | PLWH_{AA} | $p$ value$^c$ |
|-------|----------|------------|------------|--------------|
|       | n | Mean (standard deviation) | n | Mean (standard deviation) | n | Mean (standard deviation) |
| HDL   | Pooled | 3095 | 41.35 (13.22) | 3189 | 48.36 (15.44) | < 0.0001 |
|       | Female | 332 | 47.51 (15.47) | 990 | 52.98 (16.57) | < 0.0001 |
|       | Male | 2763 | 40.61 (12.72) | 2199 | 46.29 (14.44) | < 0.0001 |
|       | $p$ value$^d$ | . | < 0.0001 | . | < 0.0001 | . |
| LDL   | Pooled | 2926 | 107.6 (31.20) | 3138 | 100.1 (32.99) | < 0.0001 |
|       | Female | 317 | 107.2 (29.46) | 975 | 103.6 (33.93) | 0.0689 |
|       | Male | 2609 | 107.6 (31.41) | 2163 | 98.58 (32.45) | < 0.0001 |
|       | $p$ value$^d$ | . | 0.839 | . | < 0.0001 | . |
| Triglycerides | Pooled | 3083 | 206.7 (171.7) | 3175 | 155.6 (103.3) | < 0.0001 |
|       | Female | 331 | 185.2 (141.1) | 986 | 142.3 (76.86) | < 0.0001 |
|       | Male | 2752 | 209.3 (174.8) | 2189 | 161.60 (112.8) | < 0.0001 |
|       | $p$ value$^d$ | . | 0.0045 | . | < 0.0001 | . |

$^a$PLWH_{EA}, European American sub-cohort of people living with HIV. $^b$PLWH_{AA}, African American sub-cohort of people living with HIV. $^c$p values for differences in each continuous variable by race. $^d$p values for differences in each continuous variable by gender
having sufficient statistical power (Fig. 1, bottom panel; Additional file 2: Table S5). Specifically, we identified 12 independent loci associated with HDL levels, including intronic variants in TMTC2, CYP2B6, GRM7, BARX2, IGF2BP1, CEMIP, TNFAIP8; 11 independent loci associated with LDL levels, including intronic variants in LBR, PRKG1, RCOR1, TNIP1, PRKAG2, and seven independent loci associated with triglyceride levels, including variants in SBK1, GPR156, and CPA6 (Additional file 3: Table S5). In a subgroup analysis of PLWHAA, in addition to replicating previously reported associations of APOE, TOMM40, and NECTIN2 with LDL, HERPUD1/CETP with HDL, and APOA5 with triglycerides at the genome-wide significance level, and of APOB, CELSR2, and LDLR with LDL and LPL, LIPC, and DOCK7 with triglyceride levels at $p < 10^{-5}$ (Fig. 2, top panel,
Additional file 4: Table S6), we found lipid-related loci that were unique to PLWH AA (Fig. 2, bottom panel, Additional file 5: Table S7). Specifically, we identified 18 independent HIV-specific loci associated with HDL, 11 with LDL, and seven with triglyceride levels in PLWH AA at \( p < 5 \times 10^{-5} \), including intergenic variants in \( CPA6 \), previously associated with total cholesterol [57] and T2D [58] in individuals of African ancestry, and \( PRKG1 \) linked to body composition [59]. Lastly, we provide further evidence suggesting sex-specific effects of lipid-related SNPs. While none of these associations achieved genome-wide statistical significance (Additional file 6: Table S8), as a group, the corresponding genes were enriched in the visceral fat deposits and the metabolic syndrome pathways using BioCarta as implemented in Enrichr [49].

**Gene set enrichment analysis**

Gene set enrichment analysis was performed using genes containing HIV-specific susceptibility loci identified through GWAS\(_{HIV}\) of HDL (599 genes), LDL (595 genes), and triglycerides (678 genes). We identified several significantly enriched terms in the UK Biobank GWAS (version 1) gene set library (Fig. 3). Several top enriched terms were associated with blood cell counts, body composition, fat measurements and distribution, hypertension, diabetes, mood changes, and behavioral risk factors, such as alcohol dependence and smoking. Several of these enriched terms were statistically significant in all three gene set enrichment analyses, i.e., using HIV-specific variants from GWAS\(_{HIV}\) of HDL, LDL, and triglycerides (Fig. 3).

**Expression quantitative trait loci**

Given the association between HIV-specific lipid-related loci and immune cell counts (Fig. 3), we compared the proportion of eSNPs among the HIV-specific SNPs with the proportion of eSNPs among all remaining SNPs in various CD14+ monocyte eQTL data sets (at basal condition, IFN-\(\gamma\)-induced, LPS-induced for 2-h, and LPS-induced for 24-h). The eSNPs were significantly enriched among the HIV-specific SNPs for HDL and LDL \(( p < 0.01)\) for all conditions except for basal condition for LDL SNPs (Additional file 1: Table S9 and Fig. S1). For triglycerides, the enrichment was significant only in the non-induced cells.

**PRS analysis**

We first tested the association of various lipid levels and risk of MI or T2D in CNICS patients with PRS for corresponding traits and diseases derived from GWAS\(_{GEN}\) (Additional file 1: Table S2) at eight different GWAS \( p \) value thresholds. We detected highly significant correlations between PRS for lipid traits (HDL, LDL, and triglycerides) and corresponding phenotypes (e.g., PRS\(_{HDL}\) and plasma HDL; Fig. 4). Furthermore, as expected, measured HDL levels were inversely correlated with PRS for LDL, triglycerides, and CAD. Measured LDL levels were positively associated with PRS for CAD and PRS for MI. T2D diagnosis was associated with higher PRS for BMI and CAD. There was a trend toward higher PRS for LDL associated with the risk of MI diagnosis.

For each lipid trait, we compared the variance explained (adjusted \( R^2 \)) by the PRS\(_{GEN} \) [31] versus
multiethnic PRS\textsubscript{HIV} separately in PLWH\textsubscript{EA} and PLWH\textsubscript{AA} (Fig. 5). PRS\textsubscript{GEN} explained up to 6% of the genetic variance in PLWH\textsubscript{EA} (Fig. 5a, x-axis), but only up to 4% in the PLWH\textsubscript{AA} sub-cohort (Fig. 5b, x-axis). Among the lipid traits, the largest variance explained by PRS\textsubscript{GEN} was for HDL in PLWH\textsubscript{EA} and for LDL in PLWH\textsubscript{AA}, whereas the smallest was for triglycerides. Moreover, in PLWH\textsubscript{AA}, using the multiethnic PRS\textsubscript{HIV} increased the $R^2$ for LDL across all $p$ value thresholds and for HDL, especially when variants with more stringent $p$ values were included. In PLWH\textsubscript{EA}, PRS\textsubscript{HIV} performed as well as PRS\textsubscript{GEN}, with the highest $R^2$ recorded for HDL across most of $p$ value thresholds (Fig. 5).

Lastly, to determine the predictive value of different PRS in the presence of HIV infection, we estimated the risk of T2D and MI among PLWH with the highest PRS\textsubscript{GEN} (PRS\textsubscript{GEN} for T2D and PRS\textsubscript{GEN} for MI, respectively) or the highest multiethnic PRS\textsubscript{HIV} (PRS\textsubscript{HIV} for T2D and PRS\textsubscript{HIV} for MI, respectively). For T2D, PLWH\textsubscript{EA} at the top 5% of PRS\textsubscript{GEN,T2D} had an up to 2.14-fold increased risk depending on the GWAS $p$ value threshold used for derivation compared to the remaining 95% (Fig. 6, Additional file 7: Table S10). Stratification based on PRS\textsubscript{GEN} for T2D was unable to distinguish PLWH\textsubscript{AA} at higher risk of T2D. However, PLWH\textsubscript{AA} at the top 5% of the multiethnic PRS\textsubscript{HIV,T2D} had an up to 2.35-fold increased risk (Additional file 7: Table S10). Importantly, although PRS\textsubscript{GEN} for MI was not predictive of MI risk in PLWH\textsubscript{AA}, patients at the top 5–30% of the multiethnic PRS\textsubscript{HIV} for MI had a consistently increased
Discussion

In the largest genetic study in an ethnically diverse cohort of PLWH to date, we confirmed the role of numerous susceptibility loci previously associated with lipid levels in the general population of European descent [31]. In addition, we detected variants uniquely associated with lipid traits in GWAS HIV and not in the large well-powered GWAS GEN of 188,577 individuals [31]. These HIV-specific loci were particularly enriched in eQTLs in basal and induced monocytes and associated with blood cell counts, body metabolism, mood disorders, and predisposition to risky behaviors. Lastly, we demonstrated a predictive value of PRS derived from GWAS HIV in stratifying PLWH EA to distinguish individuals at a higher risk of developing T2D, while top percentiles of multiethnic PRS derived from GWAS HIV and not PRS GEN were associated with increased risk of T2D or MI in PLWH AA.

Earlier targeted genotyping studies in general population have reported the role of genome-wide significant susceptibility loci in cardiometabolic traits in PLWH. Specifically, GWAS-validated SNPs in the APOE, APOB, LDLR, and other genes have been demonstrated to contribute to dyslipidemia in the presence of HIV infection [60]. Also, several SNPs and genetic regions common across HIV-positive and HIV-negative women have been detected in association with carotid artery intima-media thickness, a subclinical marker of atherosclerosis [61]. In a series of unbiased GWAS of lipid traits, we confirmed genetic association with previously reported variants in several apolipoprotein-coding genes (APOE, APOC1, APOC2, APOC4, and APOA5), CETP, LPL, BUD13, AQP9, and CELSR2, among many others (Fig. 1, Additional file 2: Table S4).

Additionally, we detected numerous loci that were associated with lipid traits in the PLWH EA, but showed no significant signal in the large lipid GWAS conducted in a cohort of European ancestry [31] (Fig. 1, Additional file 3: Table S5). A few small GWAS studies performed in HIV-infected cohorts have identified loci associated with carotid atherosclerosis [26], subcutaneous adipose tissue volume [25], and fat loss [24]. In our study, many of the lipid-related susceptibility loci identified in GWAS HIV were also linked by previous studies to HIV viral load [62], susceptibility [63], control [64], smoking behavior [65–67], alcohol dependence [64, 65, 68–70], and cannabis dependence [71–73], more common in PLWH than in individuals without HIV, suggesting the contribution of additional genetic variants associated with HIV infection and adverse lifestyle behaviors to dyslipidemia in this population. Importantly, HIV-specific lipid-related variants were also significantly enriched among the loci associated with blood cell counts, body
composition, lifestyle risk factors (alcohol dependence and smoking), and mood disorders (Fig. 3). These findings are consistent with previous reports showing a positive correlation between lymphocyte count and LDL cholesterol levels [74]. Moreover, a shared link has been established between CAD risk and reticuloocyte indices, where increased hemolysis associated with high reticuloocyte counts may lead to oxidative stress and inflammation [75]. Additionally, a longitudinal relationship of depressive and anxiety symptoms with dyslipidemia and abdominal obesity has been reported [76], which can be partially explained by chronic low-grade inflammation and smoking [77]. While HIV-associated chronic inflammation has long been considered a risk factor of CVD in PLWH [78], our findings suggest that genetic variants may lead to further immune perturbations that contribute to cardiometabolic risk, especially in the presence of HIV infection. Furthermore, when we screened eQTLs in basal and induced CD14+ monocytes of healthy volunteers of European ancestry [52] for the presence of HIV-specific loci, we found significant enrichment for lipid-associated variants, further supporting a functional role of these loci in gene expression regulation of dyslipidemia in the presence of HIV infection. Validation in an independent cohort will be needed to verify the effect of HIV-specific loci on cardiometabolic diseases.

We conducted subgroup analyses to identify lipid-related genetic loci that are unique to PLWHA (Fig. 2) or act in a sex-specific manner (Additional file 6: Table S8). While none of the associations reached genome-wide significance, we identified a number of genes that have been previously associated with total cholesterol [57] and T2D [58] in individuals of African ancestry, or linked to body composition [59]. The sex-specific genes as a group were enriched in the visceral fat deposit and the metabolic pathways. Additional analyses will be required to dissect the ancestry and sex-specific effects of these variants on metabolic traits in the presence of HIV infection.

Given the polygenic nature of CAD and its numerous risk factors, PRS-based assessment of the genetic burden across multiple susceptibility loci has demonstrated greater predictive value for disease risk and drug response than individual variants [33]. A recent study in a non-HIV cohort has shown that the CAD risk associated with a high polygenic load for lipid-increasing variants was proportional to their impact on lipid levels [79]. We showed a significant correlation of PRS for lipid traits, T2D, and MI generated based on the large European GWAS (Additional file 1: Table S2) with respective phenotypes in PLWH (Fig. 4). Similar to the general population, in PLWH, we observed a positive association of PRS for CAD and PRS for MI with LDL and a negative association with HDL. Our results suggest that lipid PRS could point to modifiable risk factors in the presence of HIV infection, providing additional guidance for clinical application.

However, the variance explained by PRS derived from general (predominantly European) populations in PLWHEA was > 30% lower than that explained in PLWHA (~ 6% vs. < 4%). This finding is consistent with previous studies showing that PRS calculated using effect estimates from European GWAS were not generalizable to the African ancestry population [80]. Therefore, we calculated a multiethnic PRS, shown to significantly improve disease prediction accuracy in a non-European cohort [56], by applying weights in both EA and AA GWAS in CNICS using tenfold cross-validation. Multiethnic PRSoutperformed PRS in PLWHA, especially for HDL, but not in PLWHEA (Fig. 5).

Of note, stratification based on PRS for T2D was able to distinguish PLWH that were at a higher risk of T2D, with EA at the top 5% having a more than twofold increased risk; the impact of PRS for T2D on T2D risk in AA was less obvious (Fig. 6; Additional file 7: Table S10). A 2.75-fold increased risk of T2D in individuals of European ancestry at the top 5% of PRS for T2D has been previously reported [21]. However, the multiethnic PRS for T2D significantly improved T2D risk stratification in AA, but not in EA PLWH (Fig. 6).

In addition, while PRS for MI was unable to significantly stratify MI risk in either ethnic subgroup, multiethnic PRS demonstrated over a 3-fold increased risk in PLWHA. Multiethnic PRS for MI largely unchanged the disease risk prediction in PLWHEA. In a much larger European ancestry non-HIV cohort, a 1 standard deviation higher PRS is associated with a 33% increased risk of incident MI in participants without CAD [81]. Taken together, our findings suggest that, while the large GWAS in ethnically and racially diverse cohorts should substantially contribute to the accuracy of PRS prediction in PLWH, in the absence of such studies, multiethnic scores are feasible alternatives to identify at-risk individuals. Given that medications and intensive lifestyle interventions prevent or postpone the progression to T2D and MI [82, 83], ascertainment of PLWH with high PRS may provide an opportunity to target these interventions with increased precision.

This study has some limitations. In the general population-based cohorts used in our analyses, HIV infection-related information may not have been collected or considered during recruitment or analysis. Therefore, it is possible to have an unknown number of PLWH in these cohorts. However, the rate of HIV infection in the US population is relatively low (~ 1 in 300), and inclusion of such individuals in our analyses would bias the results toward the null. We controlled for ART presence or absence and made no distinctions across
ART regimens. A thorough investigation of the effects of ART on lipids, which is a rapidly evolving field, is a big task and beyond the scope of the present analysis. Future investigations may be able to refine some of the work done in our study. We performed analyses of PRS for BMI but did not analyze the observed BMI. Many factors are associated with BMI among PLWH, including body morphology disorders and lifestyle, and fully analyzing these characteristics was beyond the scope of this study. Future work should elucidate relationships with the observed BMI. Additionally, we used the same cohort for multiethnic PRS derivation and validation; however, we do not expect over-fitting to be a concern given the small number of mixing weights optimized (up to 2) relative to the target sample size (> 3000) and given our use of adjusted $R^2$ as the evaluation metric, similar to previously reported analyses [56]. In order to minimize the possibility of an inflated $R^2$ prediction due to shared population stratification or familial/distant relatedness [84], we used ancestry-adjusted regression coefficients for PRS computation and ten-fold cross-validation. Despite being the largest genetic study reported in PLWH, the number of MI cases was too small to provide sufficient statistical power to assess the clinical impact of PRS. Nevertheless, we were able to demonstrate that the use of multiethnic PRS in PLWH outperformed PRS derived in largely European populations, especially for PLWH AA. Going forward, meta-analyses of PLWH cohorts should allow for validation of our findings and help assess the clinical impact of the genetic burden on disease risk.

Conclusions
In summary, we demonstrated that in addition to genetic loci in the lipid metabolism genes previously linked to dyslipidemia and other CAD-related risks in the general population, there are other genetic factors that can impact lipid levels by further enhancing inflammation and predisposing to mood disorders and risky behaviors, thereby contributing to dyslipidemia in the presence of HIV infection. Comprehensive polygenic risk profiling identified PLWH to be at a several-fold increased risk of T2D or MI, which may help increase the precision of ascertaining those at high risk for targeted interventions.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s12916-020-01762-z.

Additional file 1: Table S1. A summary of genotyping and imputation platforms utilized to generate the final study data set. Table S2. Description of previously published genome-wide association studies (GWAS) in European populations that were used in polygenic risk score analyses. Table S3. Number of variants used to derive various polygenic risk scores for different traits and diseases. Table S9. Number of expression quantitative trait loci controlled by HIV-specific loci in CD14+ monocytes from Fairfax et al. [51]. Fig. S1. Fraction of expression quantitative trait loci among the HIV-specific loci (see the Methods section for details) compared to all other loci. HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglycerides; SNP, single nucleotide polymorphism; CD14, CD14+ monocytes at baseline; INF, activated CD14+ monocytes following induction with interferon-γ; LPS2, activated CD14+ monocytes following a 2-h induction with lipopolysaccharide; LPS2A, activated CD14+ monocytes following a 24-h induction with lipopolysaccharide; SNP, single nucleotide polymorphism.

Additional file 2: Table S4. List of variants significantly associated with lipid levels in European Americans in the CNICS HIV cohort (p < E-5) and in Willer et al. [31] (p < 0.05).

Additional file 3: Table S5. List of variants significantly associated with lipid levels in European Americans in the CNICS HIV cohort (p < E-5), non-significant in Willer et al. [31] (p > 0.05) and with no overlap between 99% confidence intervals for beta coefficients between CNICS and Willer et al.

Additional file 4: Table S6. List of variants significantly associated with lipid levels in African Americans in the CNICS HIV cohort (p < E-5) and in Willer et al. [31] (p < 0.05).

Additional file 5: Table S7. List of variants significantly associated with lipid levels in African Americans in the CNICS HIV cohort (p < E-5), non-significant in Willer et al. [31] (p > 0.05) and with no overlap between 99% confidence intervals for the beta coefficients between CNICS and Willer et al.

Additional file 6: Table S8. List of variants significantly associated with lipid levels in European American females (p < E-5) but not in males (p > 0.05) in the CNICS cohort with no overlap between 99% confidence intervals for beta coefficients between females and males.

Additional file 7: Table S10. Prevalence and clinical impact of high European-based and multiethnic polygenic risk scores for type-2 diabetes (T2D) and myocardial infarction (MI) in people living with HIV.

Abbreviations
AA: African American; AIDS: Acquired immune deficiency syndrome; ART: Antiretroviral therapy; BMI: Body mass index; CAD: Coronary artery disease; CNICS: Centers for AIDS Research Network of Integrated Clinical Systems; CVD: Cardiovascular disease; EA: European American; eQTL: Expression quantitative trait locus; FDR: False discovery rate; GWAS: Genome-wide association study; HDL: High-density lipoprotein cholesterol; HIV: Human immunodeficiency virus; IRB: Institutional Review Board; KEGG: Kyoto Encyclopedia of Genes and Genomes; LD: Linkage disequilibrium; LDL: Low-density lipoprotein cholesterol; MEGA: Multiethnic Global Array; MI: Myocardial infarction; PCA: Principal components analysis; PLWH: People living with HIV; PLWH AA: African American sub-cohort of people living with HIV; PLWH EA: European American sub-cohort of people living with HIV; PRS: Polygenic risk score; QC: Quality control; SNP: Single nucleotide polymorphism; T2D: Type 2 diabetes

Acknowledgements
The authors thank the study participants from the Centers for AIDS Research Network of Integrated Clinical Systems (CNICS).

Authors’ contributions
AS and IP drafted and edited the manuscript. AS, IP, HC, KH, and CML processed, analyzed, visualized, and interpreted the data. HMC, PKC, BMW, RMN, SRW, JW-N, JAD, WJL, MMK, MSS, AWilling, JEE, WCM, PWH, RDM, AWebel, and KHM participated in the cohort setup and data collection and sample processing, assisted with data acquisition and interpretation, and provided substantive comments and edits for the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the National Heart, Lung, And Blood Institute [R01HL125027 to IP and HMC and R01HL126538 to JD], National Human Genome Research Institute [R01HG010649 to IP and HMC], National Institute on Drug Abuse [R01DA047045 to HMC and IP]. Additional support came
from the National Institute of Allergy and Infectious Diseases [R24 AI067039 to MSS, P30 AI027757, and P30 AI027767 to MSS]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
The datasets generated and analyzed during the current study are available from the authors on reasonable request.

Ethics approval and consent to participate
Institutional Review Boards (IRBs) at each institution have approved CNICS research activities, and a written informed consent was obtained from all participants. The ethics approval of data analysis of de-identified subjects conducted during the current study was waived by the IRB at the Icahn School of Medicine at Mount Sinai, New York, New York.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America. 2Department of Medicine, University of California San Diego, San Diego, CA, United States of America. 3Department of Medicine, University of Washington School of Medicine, Seattle, WA, United States of America. 4Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, United States of America. 5Department of Medicine, University of Washington School of Medicine, Seattle, WA, United States of America. 6Center for AIDS Research, University of Washington, Seattle, WA, United States of America. 7School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States of America. 8Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, United States of America. 9Department of Medicine, University of California San Diego, San Diego, CA, United States of America. 10Division of Experimental Medicine, University of California San Francisco, San Francisco, CA, United States of America. 11Department of Medicine, Johns Hopkins University, Baltimore, MD, United States of America. 12Department of Epidemiology, Johns Hopkins University, Baltimore, MD, United States of America. 13Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, United States of America. 14The Fenway Institute at Fenway Health, Boston, MA, United States of America.

Received: 9 March 2020 Accepted: 24 August 2020
Published online: 28 October 2020

References
1. Global HIV and AIDS statistics - 2018 fact sheet [http://www.unaids.org/en/resources/fact-sheet].
2. Van Epps P, Kalayjian RC. Human immunodeficiency virus and aging in the era of effective antiretroviral therapy. Infect Dis Clin N Am. 2017;31(4):791–810.
3. Smith CJ, Ryom L, Weber R, Morlat P, Pradier C, Reiss P, Kowalska JD, de Wit S, Law M, el Sadr W, et al. Trends in underlying causes of death in people diagnosed with HIV in the era of highly active antiretroviral therapy compared with the general population: an analysis of a national observational cohort. Lancet Public Health. 2017;2(1):e35–e10.
4. Croxford S, Kitching A, Desai S, Kall M, Edelstein M, Skingsley A, Burns F, Copas A, Brown AE, Sullivan AK, et al. Genetic and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707–13.
5. competencies taught us about the biology of obesity and its complications. Lancet Diabetes Endocrinol. 2018;6(3):223–36.
6. Gonzalez-Muniesa P, Martinez-Gonzalez MA, Hu FB, Despres JP, Matsuzawa Y, Loos RIF, Moreno LA, Bray GA, Martinez JA. Obesity. Nat Rev Dis Primers. 2017;3:17034.
7. Ingelsson E, McCarthy MI. Human genetics of obesity and type 2 diabetes mellitus. Curr Genom Precis Med. 2018;11(1):e002090.
8. Kathiresan S, Silvestra D. Genetics of human cardiovascular disease. Cell. 2012;148(6):1242–57.
9. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet. 2018;19(2):110–24.
10. Torkamani A, Winerenger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581–90.
11. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748–52.
12. Ripatti S, Tikkanen E, Ortho-Melander M, Havašinla AS, Sillerud K, Sharma A, Guiducci C, Perola M, Julia A, Similä J, et al. A multic locus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010;376(9750):1393–400.
13. Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, Wiklund O, Helgeja RA, Fjel D, Felseiche J, et al. Familial hypercholesterolemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. Eur Heart J. 2013;34(45):3478–3491a.
14. Abu-Husn NS, Manickam C, Jones LK, Wright EA, Hartzell DN, Gonzalez-Jauregui C, O’Dushlaine C, Leader JB, Lester Kirchner H, Lindbream ME, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. Science. 2016;354(6319):aaal0070.
15. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Chi D, Shah J, Natarajan P, Lander ES, Roberts R, et al. Genome-wide polygenic scores for blood lipids and amyotrophic lateral sclerosis. Neurobiol Aging. 2018;67:202–e12.
16. Semenchenko E, Shrestha S, Yu Y, Gotuzzo E, Lobitz W, et al. Genetic epidemiology of secondary prevention trials. Lancet. 2015;385(9984):2264–71.
17. Copas A, Brown AE, Sullivan AK, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet. 2015;385(9984):2264–71.
18. Chen X, Yazar M, Pietilä M, Magnusson PK, Fang F. Polygenic link between blood lipids and amyotrophic lateral sclerosis. Neurobiol Aging. 2018;67:202–e12.
19. Uyttayakul S, Oudot-Mellah T, Nakayama EE, Tengtrakulcharoen P, Guergnon J, Delfassy JF, Khumsam P, Sangsajja C, Likanonsakul S, Theodorou I, et al. Genome-wide association study of carotid atherosclerosis in HIV-infected men. Pharmacogenet Genomics. 2011;21(12):799–807.
20. Shrestha S, Irvin MR, Taylor KD, Wiener HW, Pajewski NM, Haritunians T, et al. Genetics linked to energy metabolism and immunoregulatory mechanisms are associated with subcutaneous adipose tissue distribution in HIV-infected men. Pharmacogenet Genomics. 2011;21(12):799–807.
21. Shrestha S, Irvin MR, Taylor KD, Wiener HW, Pajewski NM, Haritunians T, Delaney JA, Schambelan M, Polak JF, Arnett DK, et al. A genome-wide association study of carotid atherosclerosis in HIV-infected men. AIDS. 2010;24(4):583–92.
background, traditional risk factors, and HIV-related factors to coronary artery disease events in HIV-positive persons. Clin Infect Dis. 2013;57(1):112–21.

28. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90.

29. Nilkay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, et al. A comprehensive 1,000 genome-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121–30.

30. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, Zeng L, Ntalla I, Lai FP, Hopewell JC, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017;49(9):1385–91.

31. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–83.

32. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206.

33. Natajaran P, Young R, Sztitkol NO, Padmanabhan A, Baber U, Mehran R, Sartori S, Fuster V, Reilly DF, Butterworth A, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. Circulation. 2017;135(22):2091–101.

34. Khatana MM, Rodriguez B, Haubrich R, Boswell S, Mathews WC, Lederman MM, Lobor WB, Van Rompaey SE, Crane HM, Moore RD, et al. Cohort profile: the centers for AIDS research network of integrated clinical systems. Int J Epidemiol. 2008;37(3):948–55.

35. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18(6):499–502.

36. Crane HM, Kadane JB, Crane PK, Khatana MM. Diabetes case identification methods applied to electronic medical record systems: their use in HIV-infected patients. Curr HIV Res. 2006;4(1):97–106.

37. Crane HM, Paramsothy P, Drozd DR, Nance RM, Delaney JA, Heckbert SR, Budoff MJ, Burkeholder GA, Willing JH, Mugavero MJ, et al. Types of myocardial infarction among human immunodeficiency virus-infected individuals in the United States. JAMA Cardiol. 2017;2(3):260–7.

38. Crane HM, Heckbert SR, Drozd DR, Budoff MJ, Delaney JA, Rodriguez C, Paramsothy P, Lobor WB, Burkeholder G, Willing JH, et al. Lessons learned from the design and implementation of myocardial infarction adjudication tailored for HIV clinical cohorts. Am J Epidemiol. 2014;179(5):996–1005.

39. Thygesen K, Alpert JS, White HD, Jaffe AS, Apple FS, Galvani M, Katus HA, Newby LK, Ravkilde J, Chaitman B, et al. Universal definition of myocardial infarction. Circulation. 2007;116(22):2563–54.

40. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD. Third universal definition of myocardial infarction. Eur Heart J. 2012;33(20):2551–67.

41. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, EY, Levy S, McGue M, et al. Next-generation genotype imputation service for large- and ultra-large-scale functional genomics studies. Bioinformatics. 2012;28(19):2543–40.

42. Schurawer GI, Galecki AT, Blasi L. Sample size and the transition to phase-3 dataset. Circ Cardiovasc Genet. 2009;2(6):621–90.

43. Jin Y, Schaffer AA, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90.

44. Kukesh MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KW, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90–7.

45. Sudlow C, Gallagher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

46. Schwarzer G. Meta: an R package for meta-analysis. R news. 2007;7(4):40–5.

47. Taylor KC, Carty CL, Dumitrescu L, Buzkova P, Cole SA, Hindorff L, Schumacher FR, Wilkens LR, Shoget RV, Quijano RM, et al. Investigation of gene-by-sex interactions for lipid traits in diverse populations from the population architecture using genomics and epidemiology study. BMC Med. 2013;14:33.
analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use. Biol Psychiatry. 2019;85(11):946–55.

67. Park SL, Carmella SG, Chen M, Patel Y, Stram DO, Haiman CA, Le Marchand L, Hecht SS. Mercapturic acids derived from the toxicants acrolein and crotonaldehyde in the urine of cigarette smokers from five ethnic groups with differing risks for lung cancer. PLoS One. 2015;10(6):e0124841.

68. Kapoor M, Wang JC, Wetherill L, Le N, Bertelsen S, Hinrichs AL, Budde J, Agrawal A, Almasy L, Bucholz K, et al. Genome-wide survival analysis of age at onset of alcohol dependence in extended high-risk COGA families. Drug Alcohol Depend. 2014;142:56–62.

69. Evangelou E, Gao H, Chu C, Ntzorras G, Blakeley P, Butts AR, Papke L, Suzuki H, Koskeridis F, Yioskas AM, et al. New alcohol-related genes suggest shared genetic mechanisms with neuropsychiatric disorders. Nat Hum Behav. 2019;3(9):950–61.

70. de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, Schwander K, Kraja AT, Guo X, Franceschini N, et al. Multiancestry genome-wide association study of lipid levels incorporating gene-alcohol interactions. Am J Epidemiol. 2019;188(6):1033–54.

71. Hubel C, Gaspar HA, Coleman JR, Hanscombe KB, Purves K, Prokopenko I, Graf M, Ngwa JS, Workalahu M, Consortium AWGotPG, et al. Genetic correlations of psychiatric traits with body composition and glycemic traits are sex- and age-dependent. Nat Commun. 2019;10(1):5765.

72. Kranzler HR, Zhou H, Kember RL, Smith RV, Justice AC, Damrauer S, Tsao PS, Klarin D, Baras A, Reid J, et al. Author correction: genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019;10(1):4050.

73. Sanchez-Roige S, Fontanillas P, Elson SL, 23andMe Research Team, Gray JC, de Wit H, Davis LK, MacKillop J, Palmer AA. Genome-wide association study of alcohol use disorder identification test (AUDIT) scores in 20,328 research participants of European ancestry. Addict Biol. 2019;24(1):121–31.

74. Oda E. Longitudinal associations between lymphocyte count and LDL cholesterol in a health screening population. J Clin Transl Endocrinol. 2014;1(2):49–53.

75. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167(5):1415–29.e19.

76. van Reedt Dortland AK, Giltay EJ, van Veen T, Zitman FG, Penninx BW. Longitudinal relationship of depressive and anxiety symptoms with dyslipidemia and abdominal obesity. Psychosom Med. 2013;75(1):83–9.

77. van Reedt Dortland AK, Vreeburg SA, Giltay EJ, Licht CM, Vogelzangs N, van Veen T, de Geus EJ, Penninx BW, Zitman FG. The impact of stress systems and lifestyle on dyslipidemia and obesity in anxiety and depression. Psychoneuroendocrinology. 2013;38(2):209–18.

78. Barnes RP, Lacson JC, Bahrami H. HIV infection and risk of cardiovascular diseases beyond coronary artery disease. Curr Atheroscler Rep. 2017;19(5):20.

79. Ripatti P, Ramo JT, Kissilovski B, Sopry CM, Vogelzangs N, van Veen T, de Geus EJ, Penninx BW, Zitman FG. The impact of stress systems and lifestyle on dyslipidemia and obesity in anxiety and depression. Psychoneuroendocrinology. 2013;38(2):209–18.

80. Newman JD, Schwartzbard AZ, Weintraub HS, Goldberg I, Berger JS. Primary prevention of cardiovascular disease in diabetes mellitus. J Am Coll Cardiol. 2017;70(7):883–93.

81. Hoskin MA, Bray GA, Hattaway K, Khare-Ranade PA, Pomeroy J, Semler LN, Weinzierl VA, Wylie-Rosett J, for the Diabetes Prevention Program Research Group. Prevention of diabetes through the lifestyle intervention: Lessons learned from the diabetes prevention program and outcomes study and its translation to practice. Curr Nutr Rep. 2014;3(4):364–78.

82. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 2013;14(7):507–15.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:
• fast, convenient online submission
• thorough peer review by experienced researchers in your field
• rapid publication on acceptance
• support for research data, including large and complex data types
• gold Open Access which fosters wider collaboration and increased citations
• maximum visibility for your research; over 100M website views per year

At BMC, research is always in progress.
Learn more biomedcentral.com/submissions