RESEARCH

Multiomic integration analysis identifies atherogenic metabolites mediating between novel immune genes and cardiovascular risk

Robert Carreras-Torres[1,2†], Iván Galván-Femenía[3,4†], Xavier Farré[4,5], Beatriz Cortés[4], Virginia Díez-Obrero[2,6], Anna Carreras[4], Ferran Moratalla-Navarro[2,6,7,8], Susana Iraola-Guzmán[4,5], Natalia Blay[4,5], Mireia Obón-Santacana[2,6,7], Víctor Moreno[2,6,7,8*] and Rafael de Cid[4,5*]

Abstract

Background Understanding genetic-metabolite associations has translational implications for informing cardiovascular risk assessment. Interrogating functional genetic variants enhances our understanding of disease pathogenesis and the development and optimization of targeted interventions.

Methods In this study, a total of 187 plasma metabolite levels were profiled in 4974 individuals of European ancestry of the GCAT| Genomes for Life cohort. Results of genetic analyses were meta-analysed with additional datasets, resulting in up to approximately 40,000 European individuals. Results of meta-analyses were integrated with reference gene expression panels from 58 tissues and cell types to identify predicted gene expression associated with metabolite levels. This approach was also performed for cardiovascular outcomes in three independent large European studies ($N = 700,000$) to identify predicted gene expression additionally associated with cardiovascular risk. Finally, genetically informed mediation analysis was performed to infer causal mediation in the relationship between gene expression, metabolite levels and cardiovascular risk.

Results A total of 44 genetic loci were associated with 124 metabolites. Lead genetic variants included 11 non-synonymous variants. Predicted expression of 53 fine-mapped genes was associated with 108 metabolite levels; while predicted expression of 6 of these genes was also associated with cardiovascular outcomes, highlighting a new role for regulatory gene *HCG27*. Additionally, we found that atherogenic metabolite levels mediate the associations between gene expression and cardiovascular risk. Some of these genes showed stronger associations in immune tissues, providing further evidence of the role of immune cells in increasing cardiovascular risk.

Conclusions These findings propose new gene targets that could be potential candidates for drug development aimed at lowering the risk of cardiovascular events through the modulation of blood atherogenic metabolite levels.

---

†Robert Carreras-Torres and Iván Galván-Femenía contributed equally to this work.

*Correspondence:
Víctor Moreno
v.moreno@iconcologia.net
Rafael de Cid
rdecid@igtp.cat
Full list of author information is available at the end of the article

Background

The global burden of dyslipidaemias in adult population has increased over the past 30 years [1]. This progressive increase is an alarming scenario because altered levels of blood lipids and fatty acids are a major risk factor for atherosclerotic cardiovascular disease (CVD) [2, 3]. CVDs are a leading cause of disease worldwide, with near a 30% of overall deaths, and 40% of premature deaths [4]. In this context, blood biomarkers would help to detect and manage the disease and prevent their development.

Blood metabolite levels are influenced by both germline genetic variations [5] and environmental factors, such as lifestyle habits, temperature, or long-term air pollution [6, 7]. To understand and estimate the contribution of single-nucleotide polymorphisms (SNPs), many metabolite genome-wide association studies (GWAS) have been carried out during the last decade [8–15]. These studies identified several genes that can be explored as pharmacological targets to modulate metabolite levels. However, most SNP-metabolite associations lay in intronic or intergenic regions and do not provide either a clear target gene or a functional contribution of the genetic region to the associated outcome. Therefore, GWAS findings are difficult to translate into clinical interventions. Over the last decade, these challenges have been addressed integrating GWAS results with functional genomics datasets through genetic instrumental analyses, such as colocalization analyses, transcriptome-wide association studies (TWAS) and Mendelian randomization (MR) approaches [16]. A major advantage of instrumental analyses is that germline genetic associations cannot be explained by reverse causation and are less susceptible to confounding. Instrumental analyses can therefore circumvent many of the inherent limitations of traditional observational studies.

The transcriptome-wide association studies (TWAS) allow the identification of associations between predicted gene expression and metabolite levels. This instrumental approach interrogates SNPs tagging functional elements that alter expression levels to identify nearby genes causally associated with an outcome. TWAS capitalize and integrate expression quantitative trait loci (eQTL) studies and large GWAS datasets to provide associations that are not products of reverse causation [17]. In addition, TWAS results from individual tissues can be jointly analysed to leverage multi-tissue eQTLs and enlarge power to detect gene expression variation potentially altering metabolite levels [18].

Tissue transcriptome profiles reflect the average gene expression across heterogeneous cells within the analysed tissue. However, it has been observed that most cells follow a few broad transcriptional programs which can be classified into major cell types: epithelial, endothelial, mesenchymal (i.e. fibroblasts, stem cells, muscle cells), neural and immune cells [19]. Therefore, the analysis of tissue type-specific eQTLs can be very informative to identify genes altering blood metabolite levels expressed within tissues categorized by major cell type composition. Previous TWAS studies associated gene expression with lipid levels and different cardiovascular outcomes [20–24]; however, the interpretation of the results is not straightforward, as within a single locus more genes than the causal gene can be identified by pleiotropic effects (bystander genes) [25, 26]. Therefore, additional lines of evidence are needed to rule out pleiotropic effects.

In this study, we utilize genetic instrumental approaches to prioritize candidate genes across tissue categories. These genes are responsible for variations in plasma metabolites that contribute to an elevated risk of cardiovascular disease. To do this, as summarized in Fig. 1, we profiled a total of 187 unique plasma metabolite levels and 3 metabolite-ratios, including lipids particles, amino acids and fatty acids, in 4974 individuals of European ancestry of the GCAT| Genomes for Life cohort, and generated the corresponding GWAS results. Then, we generated a GWAS meta-analysis, including our study and publicly available datasets, and integrated these results with gene expression reference panels. The identified fine-mapped genes were also assessed for cardiovascular outcomes in three large European studies. Finally, for genes whose expression was associated with both metabolite levels and cardiovascular risk, we conducted MR mediation analyses. These analyses aimed to estimate the effect of gene expression on cardiovascular risk, considering both the direct effect and the effect mediated by metabolite levels. This multiomic integration approach enabled us to identify candidate genes that influence variations in blood metabolite levels and contribute to cardiovascular risk.

Methods

A detailed workflow diagram of the methods is depicted in the Additional file 1.

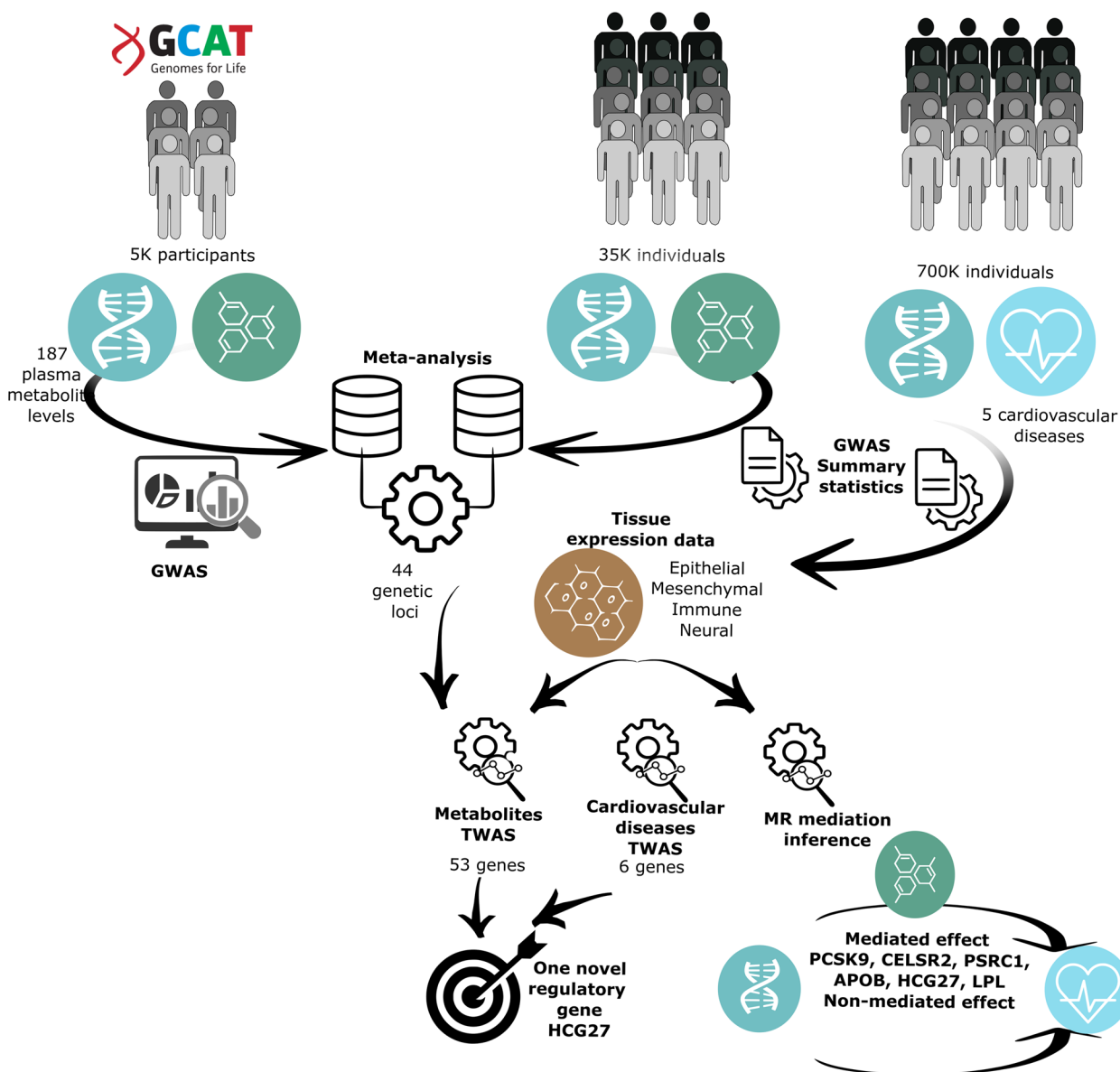Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 3 of 22

Fig. 1 Graphical abstract of the performed analyses and main results. GWAS: Genome-wide association study. TWAS: Transcriptome-wide association study. MR: Mendelian randomization

Study participants

Participants included in this study belong to the GCAT|Genomes for life cohort, a population-based cohort from Southern Europe (Catalonia, NE Spain). The GCAT cohort comprises 20,000 volunteers, recruited between 2014 and 2018, from Catalan general population, aged between 40 and 65 years old at the time of recruitment, and 59.16% women. All the participants completed a detailed, self-reported, baseline questionnaire. All participants gave their consent, and all procedures were carried out in accordance with ethical standards [27]. About 5000 participants were randomly selected from the whole cohort based on overall demographic distribution (i.e. gender, age, residence) to perform genome-wide analyses [28]. This cohort subset, named GCATcore, included 55.6% women, with a mean age of 51.0 ± 6.9 years and an average body mass index (BMI) of 27.1 ± 4.7 (Additional file 2: Table S1).

Genome data

A total of 5459 GCAT participants were genotyped using the Infinium Expanded Multi-Ethnic Genotyping Array

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 4 of 22

(MEGAEx) (ILLUMINA, San Diego, California, USA). A final dataset of 4974 GCAT participants of European ancestry that passed strict quality control were considered for genome-wide analyses, the GCATcore [28]. Genetic variants of these individuals were phased using SHAPEIT2 [29] and imputed using IMPUTE2 [30] and four reference panels: 1000 Genomes [31], UK10K [32], Genomes of the Netherlands [33] and Haplotype Reference Consortium [34]. Following the recommendations of the developers of GUIDANCE [35], the best imputation quality score from each panel was retained. After merging the imputation results of each reference panel, a total of ~20 M unique autosomal variants with minor allele frequency (MAF) > 0.001 and imputation quality score (R^2) > 0.3 were retained for subsequent analysis.

Metabolome data
A total of 188 metabolite levels were assessed in plasma samples of 4974 GCATcore participants by the Centre for Omic Sciences (COS) Joint Unit of the Universitat Rovira i Virgili-Eurecat. Details on sample processing and metabolite profiling are included in Additional file 1. In brief, each metabolite was profiled using one of these three platforms: Gas Chromatography-mass spectrometry (GCms, four batches), Liquid Chromatography-mass spectrometry (LCms, five batches) and Nuclear Magnetic Resonance Chromatography-mass spectrometry (NMRms, two batches). Redundant metabolite measures were discarded when measured using different profiling platforms. This was the case for total cholesterol levels, discarding GCms measures and keeping NMRms measures, providing a final set of 187 metabolites (Additional file 2: Table S2). Metabolites were grouped according to chemical classification provided by the Human Metabolome DataBase (HMDB) 4.0 [36]. We classified the metabolites into 4 super-classes: lipids and lipid-like molecules, organic acids, organic oxygen compounds and organoheterocyclic compounds; and 11 classes: fatty acyls, glycerophospholipids (phosphatidylcholines, lysophosphatidylcholines and phosphatidylethanolamines), glycerolipids (triglycerides), lipoprotein lipids (characteristics of lipoprotein lipids), prenols (tocopherols), sphingolipids (sphingomyelins), steroids (cholesterols), carboxylic acids (amino acids), hydroxy acids, organooxygen compounds (carbohydrates) and indoles. Additionally, we computed 3 ratio parameters: the ratio of the concentration of low-density lipoprotein (LDL) particles to high-density lipoprotein (HDL) particles [LDL/HDL], the ratio of the concentration of total lipoprotein particles to HDL particles [Total/HDL] and the ratio of leucine to isoleucine [leucine/isoleucine] (Additional file 2: Table S2). Since plasma was not always obtained in fasted conditions, analyses were adjusted correcting by chylomicrons as a closely related measure of the post-prandial phase.

Genome-wide association analysis
A linear regression was fitted for each metabolite controlling by age, sex, chylomicron levels and number of batch and the first four principal components (PC) to account for the genetic population structure of the GCAT participants. Residuals were rank-based inverse normalized and considered as the outcome for GWAS. Allele dosage for each imputed genetic variant was considered in the GWAS analysis by using PLINK2 [37].

Meta-analysis
Meta-analysis was performed using METAL software [38] and combining GWAS summary statistics from the GCAT and six additional published studies of similar ancestry, age, BMI distribution and metabolomic platforms [8–13]. Sample size, sex ratios, ancestry and age and BMI distributions are included in the Additional file 2: Table S1. A full description of the enrolling process and metabolite profiling of the additional datasets is included in the Additional file 1. The combination of the studies included up to 26,533 independent individuals and each plasma metabolite was meta-analysed by studies comprising only independent individuals using similar metabolomics platforms (GCms, LCms or NMRms). In case a metabolite was analysed by two or more studies including non-independent individuals, we considered the study with the largest sample size (Additional file 2: Table S2). Since each study performed the GWAS analysis using different transformations, meta-analyses were based on Z-scores (p-value direction) weighted by sample size using a random effects model implemented in the METAL software [38]. Assuming a standardized trait with a mean 0 and a standard deviation 1, the SNP effect (b) and standard error (se) can be estimated as suggested by Zhu et al. [39];

$$b = \frac{z}{\sqrt{2p(1-p)(n+z^2)}} \quad se = \frac{1}{\sqrt{2p(1-p)(n+z^2)}}$$

where p is the minor allele frequency (MAF) of the single-nucleotide polymorphism (SNP), n represents the sample size and Z the Metal Z-score. The threshold used to determine significant genetic variants was calculated by dividing the standard genome-wide significant threshold of $P < 5 \times 10^{-8}$ by the number of effective tests identified through a PC analysis for all analysed metabolites [13]. We found that 48 PCs explained 95% of the total variance and were considered as effective tests. Then, the significance threshold was $P < 1.04 \times 10^{-9}$.

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 5 of 22

Genetic heritability and correlation
Phenotypic variance explained by all genome-wide SNPs (heritability) was estimated through the genomic-relatedness-based restricted maximum-likelihood (GREML) approach implemented in the Genome-wide Complex Trait Analysis (GCTA) software [40]. Genetic correlations between traits were estimated using the bivariate GREML approach in GCTA [41]. Analyses were performed using the individual-level data of GCAT samples with SNPs restricted to those with MAF > 0.01. To obtain correlation results with enough statistical power, we included metabolites with nominally significant genetic heritability ($P < 0.05$). To adjust for multiple testing, we considered genetic correlations with a false discovery rate (FDR) < 0.05 as robust evidence of shared genetic contribution.

Mash analysis
To validate top SNP associations, multivariate adaptive shrinkage (mash) analysis of the meta-analysed summary statistics was performed [42]. In the presence of structured effects, mash analysis shares information across metabolites and shrinks estimates towards zero allowing the identification of associated loci. We performed this analysis using the R package mashr. As a condition to nominate significant genetic variants in each GWAS, we considered a genome-wide significance threshold of 1.04×10^{-9} in the meta-analysis and a standard local false sign rate (lfsr) < 0.05 in the mash approach.

SNP annotation
Top associated SNPs were functionally annotated and gene-mapped using the integrative web-based Functional Mapping and Annotation (FUMA) platform (https://fuma.ctglab.nl/) [43]. Overlapping top SNP regions among metabolites were considered as a single locus. Previous associations with complex phenotypes were as well assessed using FUMA. Regional plots for top associated SNPs were generated, which enables inspection of the strength of association, the extent of association signal and linkage disequilibrium (LD), and the position of findings relative to genes in the region. Plots were generated using the software LocusZoom v1.3 [44]. Measures of LD were estimated using European populations of the 1000 Genomes Project.

Transcriptome-wide association analysis for metabolites
We integrated results of GWAS meta-analyses with reference panels for gene expression under the summary-based PredXcan approach [45] to identify predicted gene expression associated with metabolite blood level. Elastic net prediction models for genetically regulated gene expression in multiple tissues from GTEx v8 were obtained from PredictDB (http://predictdb.org/) [46, 47]. In addition, we retrieved the Correlated Expression and Disease Association Research (CEDAR) dataset that included transcriptome data of six relevant blood cell types (CD4, CD8, CVD19, CD14, CD15 and platelets) [48]. CEDAR data quality control and elastic net prediction models were performed by our group and was described in a previous work [49]. For each metabolite, TWAS results were meta-analysed among tissues, using the summary-based MultiXcan approach [18], for all tissues and for different tissue groups regarding the abundance of major cell type [19]: epithelial, mesenchymal, immune or neural (Additional file 2: Table S3). In order to correct the results for multiple testing, we defined as robust evidence of candidate causal gene those results with a *P*-value lower than the significance threshold corrected by Bonferroni using the number of effective tests and the total number of tested genes ($P < 1.21 \times 10^{-8}$).

Fine-mapping in extended LD regions for metabolites
To tease out the causal gene in genetic regions with more than one associated gene due to extended LD, probabilistic fine-mapping was performed using the fine-mapping of causal gene sets (FOCUS) software [50]. FOCUS models the correlation structure across predictive models and computes posterior probabilities (PIP) for a gene to explain observed association signals at a specific locus. As a condition to nominate significant genes whose gene expression is probably causally associated with metabolite levels, we considered a significance threshold of 1.21×10^{-8} in the TWAS approach, and a credible set with a nominal confidence > 90% and with a PIP > 0.5 in the FOCUS approach.

Finally, fine-mapped genes were linked to diseases using the DisGeNET platform that integrates data from curated repositories, GWAS catalogues and scientific literature [51]. Diseases were classified into 23 categories according to the Medical Subject Headings 2022 (MeSH) database (https://meshb.nlm.nih.gov/treeView), and we tested whether these categories were overrepresented in our gene set using a fisher test. Significance threshold was corrected by Bonferroni using the number of tested categories ($P < 2.2 \times 10^{-3}$).

Transcriptome-wide association analysis of fine-mapped genes for cardiovascular outcomes
For genes which predicted gene expression was found associated with metabolite levels, we tested whether predicted gene expression was additionally associated with cardiovascular risk. To do so, we used cardiovascular-related GWAS results from large studies of European individuals (CARDIoGRAMplusC4D [52], FinnGen [53] and UK Biobank [54, 55]) (Additional file 2: Table S4).

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 6 of 22

A full description of the enrolling process of the cardiovascular datasets is included in the Additional file 1. The analysed cardiovascular outcomes included coronary artery disease (CAD, 3 datasets), myocardial infarction (MI, 3 datasets), angina pectoris (ANG, 2 datasets), atherosclerosis (ATH, 2 datasets) and atrial fibrillation (AF, 2 datasets). First, we integrated GWAS summary statistics with reference panels for gene expression under the summary-based PredXcan approach [45], and TWAS results were meta-analysed among tissues, using the summary-based MultiXcan approach [18], for all tissues and for different tissue groups regarding the abundance of major cell type [19]: epithelial, mesenchymal, immune or neural (Additional file 2: Table S3). In order to correct the results for multiple testing, we defined as robust evidence of candidate causal gene those results with a *P*-value lower than the significance threshold corrected by Bonferroni using the number of tested cardiovascular datasets (12 datasets) and the total number of tested genes (53 genes) ($P < 7.9 \times 10^{-5}$).

Mendelian randomization approach

Genetically informed mediation inference analysis was performed to make causal mediation inference about the relationship between gene expression, metabolite levels and cardiovascular risk. To do so, we apply a Mendelian randomization (MR) approach in several steps. MR is a method that uses genetic variants as instrumental variables to test for and estimate causal effects between risk factors and outcomes. The underlying principle is that genetic variants are randomly assigned at conception, akin to a natural randomized control trial, and germline genetic associations cannot be explained by reverse causation and are less susceptible to confounding [56]. The most used MR approach is the "two-sample" MR, where the genetic instrument of the putative risk factor is identified in one genetic study (the largest for the risk factor), but is subsequently evaluated for association to the outcome in a second genetic study (the largest for the outcome) [57].

As genetic instruments for gene expression, it was used the top associated expression quantitative trait loci (eQTL), not showing pleiotropic effects for nearby genes also identified in this study, from the most relevant tissue in the GTEx Consortium [46]. In the case of *CELSR2* and *PSRC1* genes, the selected eQTLs were not either associated to expression of *SORT1* or in LD ($R^2 = 0.08$; European samples) (https://ldlink.nih.gov/). However, the eQTL for *PSRC1* showed residual association with *CELSR2* expression in the liver. The genetic instruments for metabolite levels were selected from the top independent SNPs identified in this study. The strength of associations between a genetic instrument

and an exposure is reflected in the F-statistic, which is inversely related to weak instrument bias, being 10 the minimum estimation for a F-statistic to avoid bias of this nature [58]. The F-statistic was estimated as $F = (\frac{n-k-1}{k})(\frac{R^2}{1-R^2})$, where R^2 is the proportion of phenotypic variance explained by the genetic instrument, *n* is the sample size, and *k* the number of genetic variants [58]. Explained phenotypic variance for a single SNP was estimated as $R^2 = 2b^2(\mathrm{p})(1 - \mathrm{p})$, a function of effect size for the risk factor in standard deviation units (b) and minor allele frequency (p) [59]. The parameters of association of genetic instruments with CAD risk will be obtained from the cardiovascular-related GWAS results from large studies of European individuals (CARDIoGRAMplusC4D [52], FinnGen [53] and UK Biobank [54, 55]) (Additional file 2: Table S4).

Each genetic instrument provides an estimation of the exposure levels effect on outcome risk (Wald ratio: genetic effect on the exposure/ genetic effect on the outcome). In the case of eQTLs on metabolites and CAD risk, this was the main MR estimate. In the case of metabolite levels on CAD risk, the main MR estimate was the combination of SNP Wald ratios in a single causal estimation through an inverse-variance weighted MR estimator [57]. Because of the presence of pleiotropic variants can lead to biased causal effect estimates, several MR sensitivity analyses for data with potentially invalid instruments were applied. Initially, to evaluate the extent to which directional pleiotropy (non-balanced horizontal pleiotropy) may affect the effect estimate, we used the intercept test within an MR-Egger weighted linear regression approach [60]. Furthermore, the weighted median method relaying on the distribution on SNP effects were applied, which is less sensitive to SNPs with biased effect [61]. These MR estimates were obtained using the "TwoSampleMR" R package (R software).

Finally, we investigated complex networks of relationships between variables, in particular where some of the effect of gene expression on CAD risk may operate through an intermediate variable (metabolite levels). Under an instrumental approach, mediated and non-mediated effects of the exposure on the outcome can be estimated using a regression-based method. If all effects are linear without interaction terms, the non-mediated effect can be obtained, under the assumption of homogeneity of causal effects across individuals in the population, as the difference between the total effect of gene expression on CAD risk and the product of the effects of gene expression on metabolite levels and metabolite levels on CAD risk. The standard error and confidence intervals for these quantities can be estimated by bootstrapping [62].

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 7 of 22

Results

Metabolomic profile

We profiled metabolite blood levels of 4974 individuals of the GCATcore. The quality control procedure provided a final dataset of 187 metabolites and 3 ratios for subsequent analyses. We classified the metabolites into 11 classes according to the Human Metabolite Database (HMDB) 4.0 [36]; 10 fatty acyls, 74 glycerophospholipids (phosphatidylcholines, lysophosphatidylcholines and phosphatidylethanolamines), 26 glycerolipids (triglycerides), 23 lipoprotein lipids (characteristics of lipoprotein lipids), one prenol (tocopherol), 25 sphingolipids (sphingomyelins), 3 steroids (esterified, free and total cholesterol), 16 carboxylic acids (amino acids), 3 hydroxy acids, 5 organooxygen compounds (carbohydrates) and one indol (Additional file 2: Table S2). The distribution of metabolite levels is depicted in Additional file 1.

Genome-wide association analysis

The GCAT GWAS summary statistics were meta-analysed with six additional published studies (Additional file 2: Table S1), generating summary statistics comprising up to 26,533 independent individuals (Additional file 2: Table S2). Manhattan plots and quantile–quantile (QQ) plots of meta-analyses results for each metabolite are included in Additional file 1. As result, a total of 44 genetic loci were associated with 124 metabolite parameters, comprising 350 independent locus-metabolite associations. LocusZoom plots for the 350 locus-metabolite associations were included in Additional file 1. We identified 165 lead SNPs tagging 76 genes (Fig. 2A, B, Additional file 2: Table S5). We did not identify any significant genetic association for the other 66 analysed metabolites. The 44 genetic loci had been previously related to metabolite levels, and half of them were specific to metabolite classes; 8 of them were only associated with carboxylic acids, 6 with glycerophospholipids, 5 with lipoprotein lipids, 2 specific loci for sphingolipids and 1 for indoles. The other 22 loci were shared among metabolite classes (Fig. 2A, C). Two metabolite classes, prenols and organooxygen compounds, showed no genome-wide significant SNPs.

The functional annotation of the 165 identified lead SNPs revealed that there were 10 non-synonymous SNPs and 1 non-frameshift deletion variant (Table 1 and Additional file 2: Table S5). Among them, we
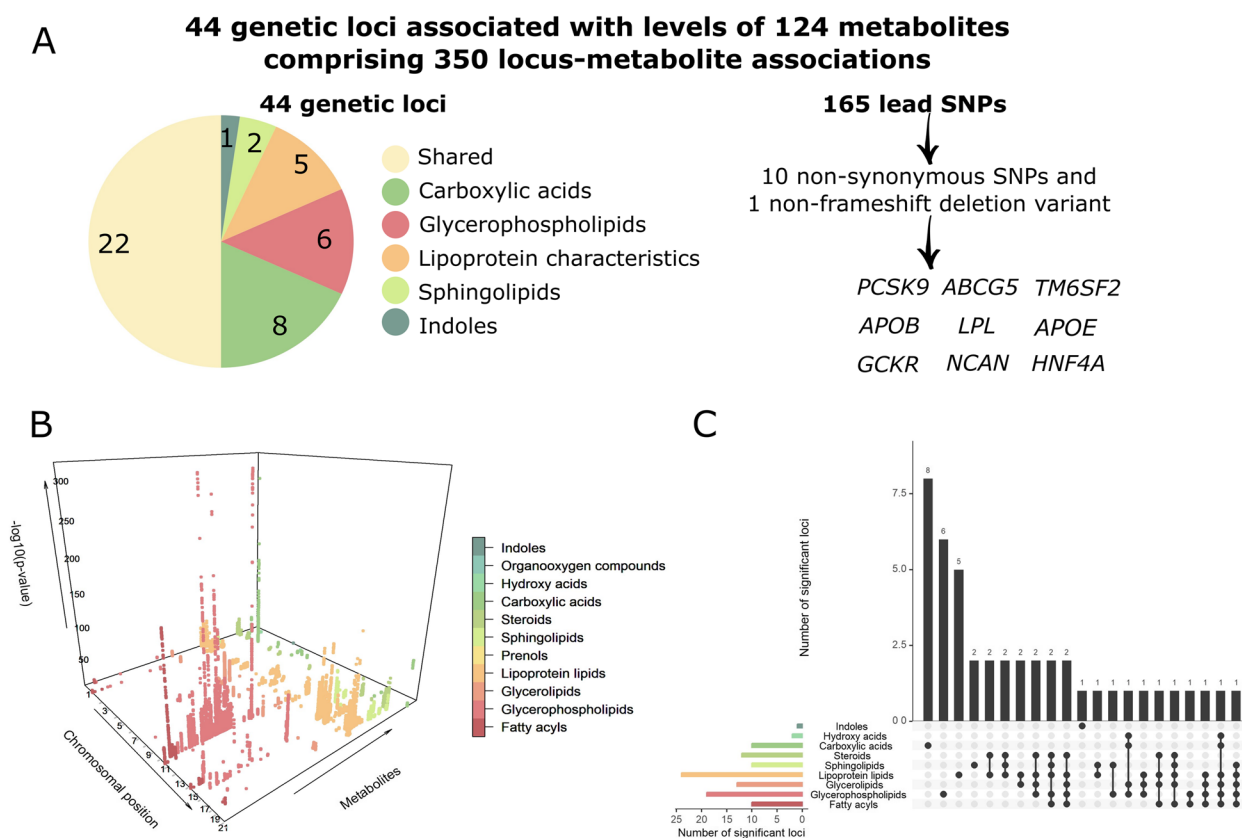

Fig. 2 Genome-wide associated variants for 190 metabolites. **A** Summary of the GWAS results. **B** 3D Manhattan plot of significant SNPs according to thresholds in meta-analyses and mashr approaches. **C** Significant loci per metabolite class

Table 1 Non-synonymous and non-frameshifting deletion variants associated with metabolite levels in GWAS meta-analyses

rs number	Chr	Bp	GCAT Freq	Eff Al	Ref Al	Class	Name	NTotal	Beta	SE	P-value	lfdr	Encode number	Symbol	Exon	Codon mutation
rs11591147	1	55,505,647	0.02	T	G	Fatty acyls	Linoleic acid	18,499	−0.298	0.039	4.E−14	5.E−47	ENSG00000169174	PCSK9	exon1	p.Arg46Leu
						Lipoprotein lipids	Large LDL	20,959	−0.553	0.037	8.E−51	7.E−48				
							Medium LDL	20,959	−0.540	0.037	2.E−48	4.E−48				
							Small LDL	20,959	−0.484	0.037	3.E−39	9.E−48				
							Cholesterol in LDL	20,959	−0.567	0.037	2.E−53	4.E−48				
						Sphingolipids	Sphingomyelin	18,447	−0.328	0.039	9.E−17	1.E−43				
						Steroids	Esterified cholesterol	18,466	−0.452	0.039	2.E−30	6.E−47				
							Free cholesterol	18,466	−0.466	0.039	2.E−32	1.E−47				
							Total cholesterol	20,959	−0.476	0.037	6.E−38	5.E−48				
rs1367117	2	21,263,900	0.29	A	G	Lipoprotein lipids	Large LDL	24,245	0.099	0.010	4.E−23	3.E−28	ENSG00000084674	APOB	exon4	p.Thr98Ile
							Cholesterol in LDL	26,531	0.105	0.010	3.E−28	6.E−29				
rs878853971	2	21,266,774	0.29	G	GGCAGCGCCA	Lipoprotein lipids	Medium LDL	24,243	0.097	0.010	2.E−22	7.E−28	ENSG00000084674	APOB	exon1	p.Leu12_Ala16delinsProAlaLeu
							Small LDL	24,243	0.088	0.010	2.E−18	8.E−28				
rs1260326	2	27,730,940	0.47	T	C	Fatty acyls	MUFA	18,509	0.078	0.010	5.E−14	1.E−07	ENSG00000084734	GCKR	exon15	p.Leu446Pro
							Omega 3	18,518	0.066	0.010	2.E−10	5.E−07				
						Glycerolipids	Triglycerides	26,519	0.078	0.009	2.E−19	2.E−13				
						Lipoprotein lipids	Small HDL	24,247	0.069	0.009	2.E−14	3.E−04				
rs6756629	2	44,065,090	0.06	A	G	Lipoprotein lipids	Large LDL	24,247	−0.120	0.019	3.E−10	8.E−13	ENSG00000138075	ABCG5	exon2	p.Arg50Cys
							Cholesterol in LDL	26,533	−0.132	0.018	4.E−13	7.E−13				
rs268	8	19,813,529	0.01	A	G	Lipoprotein lipids	Large HDL	24,243	0.264	0.039	9.E−12	2.E−21	ENSG00000175445	LPL	exon6	p.Asn318Ser

Table 1 (continued)

rs number	Chr	Bp	GCAT Freq	Eff Al	Ref Al	Class	Name	NTotal	Beta	SE	P-value	lfdr	Encode number	Symbol	Exon	Codon mutation
rs2228603	19	19,329,924	0.05	T	C	Glycerophospholipids	Phosphatidylcholine diacyl C34:4	12,450	−0.183	0.029	5.E−10	1.E−12	ENSG00000130287	NCAN	exon3	p.Pro92Ser
rs58542926	19	19,379,549	0.06	T	C	Glycerolipids	Triglycerides	26,514	−0.165	0.019	1.E−18	2.E−32	ENSG00000213996	TM6SF2	exon6	p.Glu167Lys
						Lipoprotein lipids	Triglycerides in IDL	24,242	−0.146	0.020	8.E−14	1.E−33				
						Steroids	Total cholesterol	26,460	−0.127	0.019	1.E−11	4.E−29				
rs429358	19	45,411,941	0.10	T	C	Glycerolipids	Triglycerides	26,513	−0.128	0.014	7.E−19	3.E−40	ENSG00000130203	APOE	exon4	p.Cys130Arg
						Lipoprotein lipids	Small VLDL	24,241	−0.113	0.015	5.E−14	2.E−35				
rs7412	19	45,412,079	0.06	T	C	Fatty acyls	Linoleic acid	18,497	−0.191	0.022	4.E−18	1.E−32	ENSG00000130203	APOE	exon4	p.Arg176Cys
						Lipoprotein lipids	Triglycerides in HDL	4974	0.284	0.042	2.E−11	8.E−28				
							Cholesterol in IDL	20,955	−0.350	0.021	2.E−64	3.E−72				
							LDL	4974	−0.420	0.042	4.E−23	1.E−71				
							Large LDL	20,955	−0.439	0.021	8.E−101	5.E−114				
							Medium LDL	20,955	−0.454	0.021	5.E−108	3.E−122				
							Small LDL	20,955	−0.394	0.021	1.E−81	2.E−107				
							Cholesterol in LDL	23,241	−0.495	0.019	2.E−142	6.E−148				
							Non HDL	4974	−0.398	0.042	7.E−21	6.E−65				
							LDL/HDL	4974	−0.368	0.042	5.E−18	1.E−74				
							Total/HDL	4974	−0.340	0.042	1.E−15	1.E−64				
						Steroids	Esterified cholesterol	18,466	−0.323	0.022	8.E−49	1.E−61				
							Free cholesterol	18,466	−0.226	0.022	2.E−24	1.E−37				

Table 1 (continued)

Lead SNP							Metabolite		Meta-analysis				Mashr	Gene			
rs number	Chr	Bp	GCAT Freq	Eff Al	Ref Al		Class	Name	NTotal	Beta	SE	*P*-value	lfdr	Encode number	Symbol	Exon	Codon mutation
								Total cholesterol	23,173	−0.304	0.020	1.E−53	7.E−66				
rs1800961	20	43,042,364	0.03	T	C		Lipoprotein lipids	Cholesterol in HDL	26,529	−0.166	0.026	3.E−10	3.E−07	ENSG00000101076	HNF4A	exon4	p.Thr139Ile

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 11 of 22

identified common variants such as a non-synonymous SNPs and a non-frameshifting deletion in *APOB* gene (rs1367117-Thr98Ile in exon 4 and rs878853971-p. Leu12_Ala16delinsProAlaLeu in exon 1) related to cholesterol and lipoprotein lipid levels; one SNP in *GCKR* gene (rs1260326-Leu446Pro in exon 14) associated with fatty acyls, lipoprotein lipids and glycerolipids; and the two SNPs in *APOE* gene responsible of the APOE protein isoforms (rs429358- p.Cys130Arg and rs7412-Arg176Cys in exon 4) related to fatty acyls, cholesterol and lipoprotein lipids. We also identified less frequent variants (MAF < 0.1 in the GCAT samples) such as the rs11591147 in exon 1 of *PCSK9* gene (p.Arg46Leu), the rs6756629 in exon 2 of ABCG5 (p.Arg50Cys), the rs268 (exon 6 of LPL; p.Asn318Ser), the rs2228603 (exon 3 of NCAN; p.Pro92Ser), the rs58542926 (exon 6 of TM6SF2; p.Glu167Lys) and the rs1800961 (exon 4 of HNF4A; p.Thr139Ile) associated with fatty acyls, cholesterol, lipoprotein lipids, phosphatidylcholine and sphingomyelin levels (Table 1).

Genetic heritability and correlations

SNP-based heritability estimates observed in this study (Additional file 2: Table S2) were similar than estimates observed in previous studies [63]. A total of 47 out of 190 metabolites levels and ratios showed a significant SNP-based heritability ($P < 0.05$) [$h^2 = 0.59 - 0.12$] (Additional file 2: Table S2). The metabolites showing significant genetic heritability belonged to the metabolite classes of glycerophospholipids (15 phosphatidylcholines, 2 lysophosphatidylcholines and 2 phosphatidylethanolamines), and glycerolipids (16) and lipoprotein particles mainly related with very low-density lipoproteins (VLDL) (5) (Additional file 2: Table S2). Genetic correlations were performed among the 47 metabolites with significant heritability, and a total of 40 genetic correlations were observed (FDR < 0.05), involving 15 glycerolipids, 9 glycerophospholipids and the 5 lipoprotein particles. The 85% of correlations were among metabolites of the same class (Additional file 2: Table S6).

Transcriptome-wide association analyses

We integrated GWAS meta-analyses results with reference panels for gene expression from 58 tissues and cell types. A total of 2537 gene expression-metabolite associations were identified in the TWAS meta-analyses results including all tissues (Fig. 3A), while additional 1807 gene expression-metabolite associations were identified in the meta-analyses of tissue categories (data not shown). The fine-mapping FOCUS approach validated 53 genes as associated with 108 metabolites, comprising 196 gene expression-metabolite associations (Table 2, Fig. 3B). Manhattan plots of meta-analyses results including all

tissues, and highlighting fine-mapped genes, for each metabolite were included in Additional file 1.

The 95% of fine-mapped gene expression-metabolite associations (187) identified genes potentially explaining previously observed locus-metabolite associations, from which the 74% (139) comprised the closest gene of the associated SNP (Fig. 3B, Additional file 2: Table S6). The other 5% of gene-metabolite associations (9) identified associations not observed in GWAS results (Fig. 3B), most of them within the 44 genetic loci and increasing the number of associated metabolites, except for *SEC22B* gene (1p12) related to serine levels and *GCSH* gene (16q23.2) to glycine levels (Additional file 2: Table S7). Among the 53 fine-mapped genes in Table 2, it is worth highlighting some results. The fine-mapped TWAS results suggest that variations in gene expression levels might be responsible for some of the associations observed in GWAS. For instance, *PCSK9, CELSR2, APOB, LPL, FADS1, LIPC, CETP* and *PLTP* on lipoprotein lipids; *PCSK9, CELSR2* and *APOB* on steroids; *APOB, MSH5* and *LPL* on glycerolipids; *FADS1* and *LIPC* on fatty acyls; *FADS1* and *CETP* on glycerophospholipids; and *CERS4* and *SPTLC3* on sphingomyelins (Table 2). Beside these observations, we identified *PSRC1* gene (1p13.3) associated with total LDL concentration, despite that GWAS results pointed to the SNP rs12740374, in the UTR3 region of *CELSR2* gene (Additional file 2: Table S5). We observed another similar case in the 2p23.3 region. The SNP rs6547692, located in an intron of *GCKR* gene, was found associated with concentration of small, medium and large VLDL particles (Additional file 2: Table S5). However, the gene whose predicted expression was found associated to these lipid traits was the C2orf16 gene (Table 2). Other results to highlight are found in 7q11.23 region. This genetic region was found associated to small VLDL concentration in the GWAS results; but TWAS results identified the gene *MLXIPL*, within the region, associated with concentration of small but also with medium, large and total VLDL, and triglycerides (Table 2). Also, in 10q24.31 region, the expression of *SCD* gene appeared related to levels of stearic acid and lysophosphatidylcholine acyl C16:1 (LPC16:1), while GWAS results pointed to *PKD2L1* gene. Finally, in the *APOE* loci, GWAS results only identified the two non-synonymous *APOE* SNPs; however, TWAS results also pointed to *PVR* and *APOC2* genes as potentially associated with lipoprotein lipids, steroids and sphingomyelin levels.

Most of the 53 fine-mapped genes were specific to a metabolite class (Fig. 3C), being the carboxylic acids and sphingolipids the metabolite classes with all the genes specific to the class (9 and 4 genes, respectively). These were followed by glycerophospholipids (11 genes out
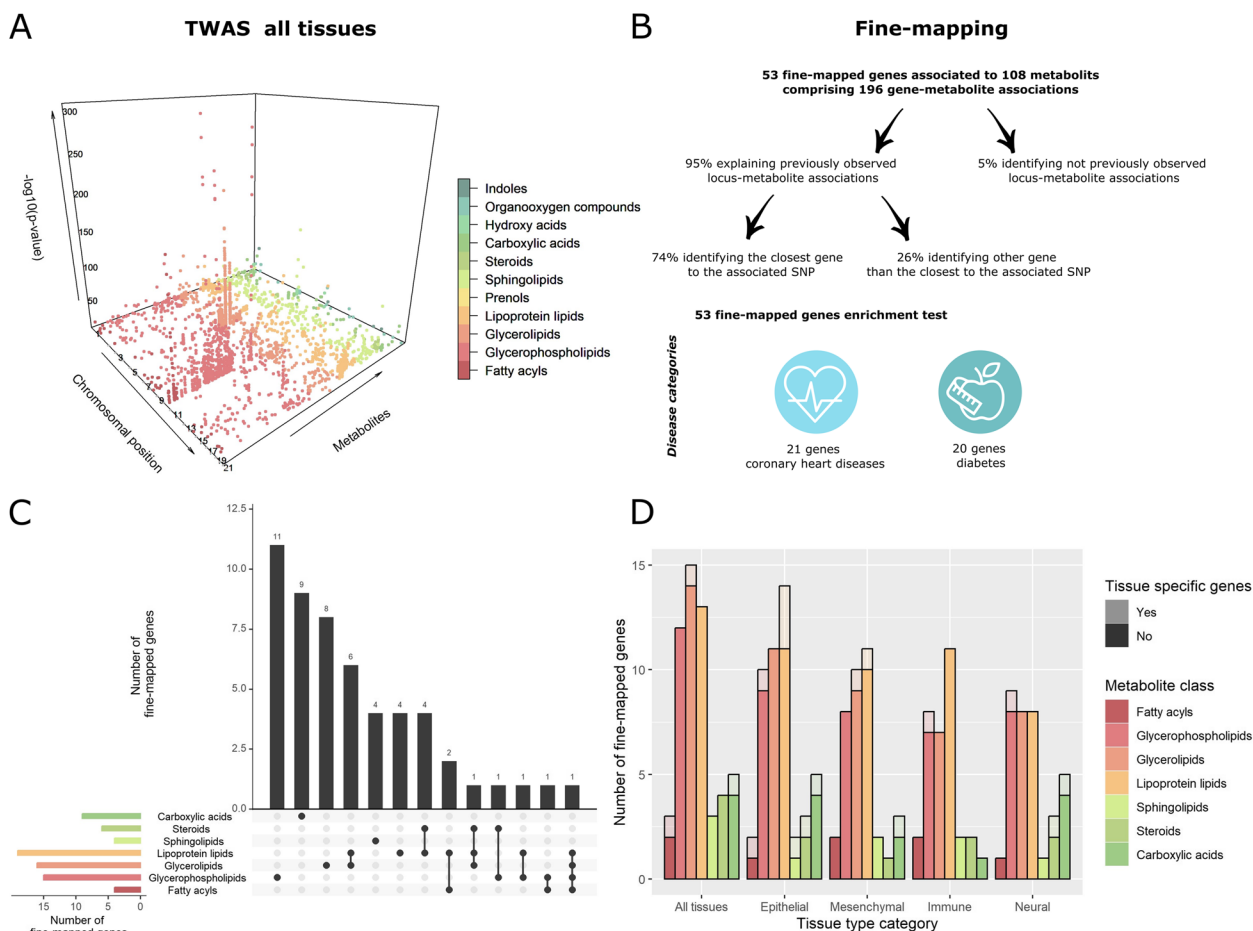
Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 12 of 22

Fig. 3 Transcription-wide associated genes for 190 metabolites. **A** 3D Manhattan plot of the 2537 gene-metabolite associations in the overall multi-tissue meta-analysis. **B** Summary of the fine-mapped 53 genes for metabolites. **C** Upset plot of the 53 fine-mapped gene-metabolite associations per metabolite class. **D** Distribution of the 53 fine-mapped genes among metabolite classes and specific cell type categories

of 15 total genes associated in this class: 11/15), glycerolipids (8/16), and lipoprotein lipids (4/18)). Steroids and fatty acyls were the classes without specific genes, 6 and 4 genes, respectively, that were mostly shared with lipoprotein lipids (Fig. 3C). On the other hand, 15 of the 53 fine-mapped genes were specific to a single tissue type category (Fig. 3D; pale colour portion of bars). For instance, the genetically predicted gene expression of *SLC7A6* was associated with lysine levels only in epithelial tissues meta-analysis, while expression of *GCSH* was related to glycine levels only in the mesenchymal tissues meta-analysis (Table 2). Finally, it is also worth to mention that some genes showed stronger effects in meta-analysis of specific tissue types compared with meta-analysis including all or other tissue types. This was the case for *PCSK9* and *SPTLC3* genes which associations with metabolite levels were stronger in immune tissue meta-analysis, and for *CELSR2* and *C2orf16* genes in epithelial tissues meta-analysis (Additional file 2: Table S7).

Cardiovascular diseases

Gene set enrichment of the 53 fine-mapped genes was tested for association with disease categories using the DisGeNET platform [51]. Results revealed that several fine-mapped genes were associated with the following DisGeNET terms: coronary heart disease (21 genes, FDR $= 3 \times 10^{-9}$), coronary artery disease (20 genes, FDR $= 9 \times 10^{-8}$) and cardiovascular diseases (19 genes, FDR $= 6 \times 10^{-7}$), while enrichment was also observed for diabetes (20 genes, FDR $= 9 \times 10^{-6}$), and diabetes mellitus (20 genes, FDR $= 1 \times 10^{-4}$) (Fig. 3B; Additional file 2: Table S8).

To identify potential causal genes whose expression contributes to cardiovascular risk, we tested for association the predicted expression of the 53 fine-mapped genes with a set of cardiovascular outcomes. The tested outcomes were coronary artery disease (CAD), myocardial infraction (MI), angina pectoris (ANG), atherosclerosis (ATH) and atrial fibrillation (AF) from three

Table 2 Fine-mapped associations of predicted gene expression for metabolites and cardiovascular outcomes

Genes				Metabolites		Cardiovascular phenotypes
Name	Chr	Bp start	Bp end	Class	Name	
TRIT1	1	40,306,723	40,349,183	Glycerophospholipids	LPC16:0[n]	
PCSK9	1	55,505,221	55,530,525	Lipoprotein lipids	Cholesterol in IDL[ai], Triglycerides in IDL[ai], Large LDL[ai], Medium LDL[ai], Small LDL[ai], Cholesterol in LDL[ai]	CGC4D CAD[ai], UKBB CAD[ai], FinnGenn CAD[ai], CGC4D MI[i], FinnGenn MI[ai], FinnGenn ANG[ai], UKBB ATH[ai], FinnGenn ATH[ai]
				Steroids	Esterified Cholesterol[a], Free Cholesterol[ai], Total Cholesterol[ai]	
DOCK7	1	62,920,399	63,153,969	Glycerolipids	Triglycerides[aeimn]	
				Lipoprotein lipids	Large VLDL[eim]	
CELSR2	1	109,792,641	109,818,372	Lipoprotein lipids	Cholesterol in IDL[aei], Large LDL[aei], Medium LDL[aei], Small LDL[aei], Cholesterol in LDL[aei]	CGC4D CAD[aeimn], UKBB CAD[aeimn], FinnGenn CAD[aei], CGC4D MI[aei], UKBB MI[aei], UKBB ANG[i], FinnGenn ANG[aei], UKBB ATH[aeimn], FinnGenn ATH[aei]
				Steroids	Esterified Cholesterol[ei], Total Cholesterol[aei]	
PSRC1	1	109,822,178	109,825,808	Lipoprotein lipids	LDL[lm]	CGC4D CAD[aeimn], UKBB CAD[aeimn], FinnGenn CAD[aeimn], CGC4D MI[aeimn], UKBB MI[aeimn], UKBB ANG[eim], FinnGenn ANG[aeimn], UKBB ATH[aeimn], FinnGenn ATH[aeimn]
ZNF697	1	120,162,045	120,190,396	Carboxylic acids	Serine[n]	
SEC22B	1	120,698,575	120,719,144	Carboxylic acids	Serine[ae]	
APOB	2	21,224,301	21,266,945	Glycerolipids	Triglycerides[an]	CGC4D CAD[am], UKBB CAD[am]
				Lipoprotein lipids	Cholesterol in IDL[aemn], Triglycerides in IDL[aemn], Large LDL[aemn], Medium LDL[aemn], Small LDL[amn], Cholesterol in LDL[aemn], Small VLDL[amn]	
				Steroids	Esterified Cholesterol[aemn], Free Cholesterol[aemn], Total Cholesterol[aemn]	
PPM1G	2	27,604,061	27,632,554	Fatty acyls	MUFA[e]	
				Lipoprotein lipids	Mean diameter for VLDL[e]	
NRBP1	2	27,650,657	27,665,126	Glycerophospholipids	LPC16:1[e]	
C2orf16	2	27,799,389	27,805,588	Glycerolipids	Triglycerides[ae]	
				Lipoprotein lipids	Large VLDL[ae], Medium VLDL[e], Small VLDL[ae]	
CPS1	2	211,342,406	211,543,831	Carboxylic acids	Glycine[aeimn]	
ELOVL2	6	10,980,992	11,044,547	Glycerophospholipids	PC38:5[aei]	
MOG	6	29,624,758	29,640,149	Carboxylic acids	Tyrosine[a]	
RPP21	6	30,312,908	30,314,661	Glycerolipids	TG54:6[ae]	
PSORS1C2	6	31,105,313	31,107,127	Glycerolipids	TG50:0[aeim]	
HCG27	6	31,165,537	31,171,745	Glycerolipids	TG52:4[aem]	CGC4D CAD[m], UKBB CAD[ai], FinnGenn CAD[i], CGC4D MI[i], FinnGenn ATH[i]
HLA-B	6	31,321,649	31,324,965	Glycerolipids	TG48:0[aimn]	UKBB CAD[an]
MICA	6	31,371,356	31,383,092	Glycerophospholipids	PCe42:5[l]	
BAG6	6	31,606,805	31,620,482	Glycerolipids	TG51:2[aemn]	

Table 2 (continued)

Genes				Metabolites		Cardiovascular phenotypes
MSH5	6	31,707,725	31,732,622	Glycerolipids	TG46:0[ae], TG46:1[ae], TG46:2[ae]	UKBB CAD[i]
SKIV2L	6	31,926,857	31,937,532	Glycerolipids	TG48:2[a]	
HLA-DRB1	6	32,546,546	32,557,625	Glycerophospholipids	PCe38:5[aem]	UKBB CAD[m]; FinnGenn ATH[n]
MFSD4B	6	111,580,551	111,592,370	Carboxylic acids	Tyrosine[aen]	
MLXIPL	7	73,007,524	73,038,873	Glycerolipids	Triglycerides[aeimn]	
				Lipoprotein lipids	Large VLDL[aeimn], Medium VLDL[aeimn], Small VLDL[aeimn], Mean diameter for VLDL[aeimn]	
LPL	8	19,759,228	19,824,769	Glycerolipids	Triglycerides[ai], TG52:2[a], TG52:3[a]	CGC4D CAD[aim],; UKBB CAD[aim], CGC4D MI[i]; UKBB MI[aim], UKBB ATH[ai]
				Lipoprotein lipids	Large HDL[aim], Mean diameter for HDL[aim], Cholesterol in HDL[aim], Large VLDL[ai], Medium VLDL[ai], Small VLDL[ai], Mean diameter for VLDL[ai], Cholesterol in VLDL[ai]	
SCD	10	102,106,881	102,124,591	Fatty acyls	Stearic acid[aim]	UKBB CAD[i]
				Glycerophospholipids	LPC16:1[aim]	
TMEM258	11	61,535,973	61,560,274	Glycerophospholipids	LPC20:5[aem]	
FADS2	11	61,560,452	61,634,826	Glycerophospholipids	PC38:3[aeimn], LPC20:3[aeimn]	
FADS1	11	61,567,099	61,596,790	Fatty acyls	Omega 3[aeimn], PUFA[aeimn], Docosahexaenoic acid[aeimn], Linoleic acid[aeimn]	
				Glycerophospholipids	PC32:0[aeimn], PC32:2[aeimn], PC34:2[aeimn], PC34:3[aeimn], PC34:4[aeimn], PC36:2[aeimn], PCe36:2[aeimn], PCe36:3[aeimn], PC36:4[aeimn], PCe36:4[aeimn], PC36:5[aeimn], PCe36:5[aeimn], PC37:4[aeimn], PC38:2[aeimn], PCe38:3[aeimn], PC38:4[aeimn], PC38:5[aeimn], PC38:5[aeimn], PC38:6[aeimn], PC40:4[aeimn], PCe40:4[aeimn], PC40:5[aeimn], PC40:6[aeimn], PCe42:4[aeimn], PCe42:5[aeimn], LPC18:2[aeimn], LPC20:4[aeimn], PEe38:5[aeimn],	
				Glycerolipids	TG52:6[aeimn]	
				Lipoprotein lipids	Mean diameter for HDL[aeimn], Cholesterol in HDL[aeimn], Mean diameter for VLDL[m]	
BEST1	11	61,717,293	61,732,987	Glycerolipids	TG54:4[m]	
ZPR1	11	116,648,436	116,658,766	Glycerolipids	Triglycerides[aeimn]	FinnGenn CAD[aem], FinnGenn ATH[aem]
				Lipoprotein lipids	Large VLDL[aeimn], Medium VLDL[aeimn], Small VLDL[aeimn], Mean diameter for VLDL[aeimn]	
SYNE2	14	64,319,683	64,693,165	Sphingolipids	SM32:1[am], SM33:1[a]	
TMEM229B	14	67,913,801	68,000,456	Glycerophospholipids	PCe32:1[aeimn], PCe36:5[aeimn], PCe38:6[aeimn]	
MYZAP	15	57,884,139	57,977,562	Glycerophospholipids	PE36:4[ae]	
ALDH1A2	15	58,245,622	58,790,065	Lipoprotein lipids	Large HDL[ai], Triglycerides in IDL[ai], Mean diameter for VLDL[i]	
LIPC	15	58,702,768	58,861,151	Fatty acyls	MUFA[a], PUFA[a]	

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 15 of 22

Table 2 (continued)

Genes			Metabolites		Cardiovascular phenotypes
FAM81A	15	59,664,892	59,815,748	Lipoprotein lipids	Small HDL[a], Mean diameter for HDL[aeimn], Triglycerides in HDL[aeimn], Mean diameter for LDL[aeimn], Triglycerides in LDL[am]
NTAN1	16	15,131,710	15,149,921	Glycerophospholipids	Phosphatidyl Choline[an]
				Steroids	Free Cholesterol[n]
CETP	16	56,995,762	57,017,757	Glycerophospholipids	PC36:3[m], PC38:3[aeimn], PCe38:3[m], LPC20:3[aeimn], LPC20:4[m]	UKBB CAD[aemn]
				Glycerophospholipids	Phosphatidyl Choline[aeimn], PCe32:1[aeimn], PCe34:1[aeim], PCe34:2[aem], PCe34:3[aeim]
				Lipoprotein lipids	HDL[aeimn], Large HDL[aeimn], Medium HDL[aeimn], Small HDL[aei], Mean diameter for HDL[aeimn], Cholesterol in HDL[aeimn], Triglycerides in IDL[m], Small VLDL[aeimn], LDL/HDL[aeimn], Total/HDL[aeimn]
SLC7A6	16	68,298,433	68,335,722	Carboxylic acids	Lysine[e]
MARVELD3	16	71,660,064	71,676,017	Carboxylic acids	Tyrosine[mn]
HPR	16	72,088,522	72,111,145	Lipoprotein lipids	Triglycerides in IDL[m]	UKBB CAD[am]
GCSH	16	81,115,566	81,130,008	Carboxylic acids	Glycine[m]
SNRPD1	18	19,192,228	19,210,417	Glycerophospholipids	PC32:2[an]
CERS4	19	8,271,620	8,327,305	Sphingolipids	SM36:1[aeimn], SM36:2[aeimn], SM38:1[aen], SM38:2[aeimn]
KRI1	19	10,663,761	10,676,713	Lipoprotein lipids	Large LDL[e]
				Steroids	Total Cholesterol[e]
ATP13A1	19	19,756,007	19,774,502	Glycerolipids	Triglycerides[aeim]
				Lipoprotein lipids	Small VLDL[e]
PVR	19	45,147,098	45,166,850	Lipoprotein lipids	Cholesterol in IDL[aen], Large LDL[an], Medium LDL[aen], Small LDL[aen], Cholesterol in LDL[aen]	UKBB CAD[an], FinnGenn CAD[a], UKBB ATH[an]
				Steroids	Esterified Cholesterol[an], Total Cholesterol[an]
APOC2	19	45,445,495	45,452,822	Sphingolipids	SM42:1[e]	UKBB CAD[ai]
SPTLC3	20	12,989,627	13,147,411	Sphingolipids	SM32:0[ai], SM32:1[a], SM33:1[ai], SM35:1[ai], SM36:0[ai], SM39:1[ai], SM40:0[ai], SM43:1[ai], SM43:2[ai]
PLTP	20	44,527,399	44,540,794	Lipoprotein lipids	Large HDL[aeim], Medium HDL[em], Small HDL[aeim], Mean diameter for HDL[aeim]
DGCR9	22	19,061,632	19,061,632	Carboxylic acids	Proline[aen]

Genes were significant in the meta-analyses including [a]all, [e]epithelial, [m]mesenchymal, [i]immune and/or [n]neural tissues. Metabolite description in Table S1

PC Phosphatidylcholine diacyl C, *PCe* Phosphatidylcholine acyl-alkyl C, *LPC* Lysophosphatidylcholine acyl C, *PE* Phosphatidylethanolamine alkyl C, *TG* Triglycerol C, *SM* Sphingomyeline C, *CGC4D* CardiogramplusC4D, *UKBB* UK Biobank, *CAD* Coronary artery disease, *MI* Myocardial infraction, *ANG* Angina pectoris, *ATH* Atherosclerosis

large studies of European population (CARDIoGRAM-plusC4D [52], FinnGen [53] and UK Biobank [54, 55]) (Additional file 2: Table S4). As result, the gene expression of 15 genes were identified as associated with cardiovascular outcomes except for AF, comprising 51 gene expression-cardiovascular outcome associations (Table 2 and Additional file 2: Table S9). Most of these associations were identified in the overall multi-tissue meta-analyses results, but additional associations were identified in the meta-analyses results including only tissues within immune (7 out of 36 total associations in this category: 7/36), mesenchymal (3/24), neural (1/15) and epithelial (0/20) tissue categories (Fig. 4A). The genes associated with an outcome in at least two datasets, thus, replicating results were *PCSK9, CELSR2, PSRC1, APOB, HCG27* and *LPL* (Fig. 4A and Table 2). Interestingly, the *PCSK9* gene showed stronger effects in the association

with cardiovascular outcomes in immune tissues meta-analysis, while *CELSR2* gene in epithelial tissues meta-analysis. Additionally, the genes APOB, HCG27 and LPL appear to play roles in immune and mesenchymal tissues. This suggests that these genes might influence immune functions and the behaviour of mesenchymal cells, which are critical for the structure but also function of connective tissues (Table 2 and Additional file 2: Table S9).

Mendelian randomization approach

Genetically informed mediation analysis was performed to infer causal mediation in the relationship between gene expression of the 6 identified genes (*PCSK9, CELSR2, PSRC1, APOB, HCG27* and *LPL*), metabolite levels and cardiovascular risk. To do so, we apply a Mendelian randomization approach in several steps. First of all, we identified the metabolites, related to the 6 genes
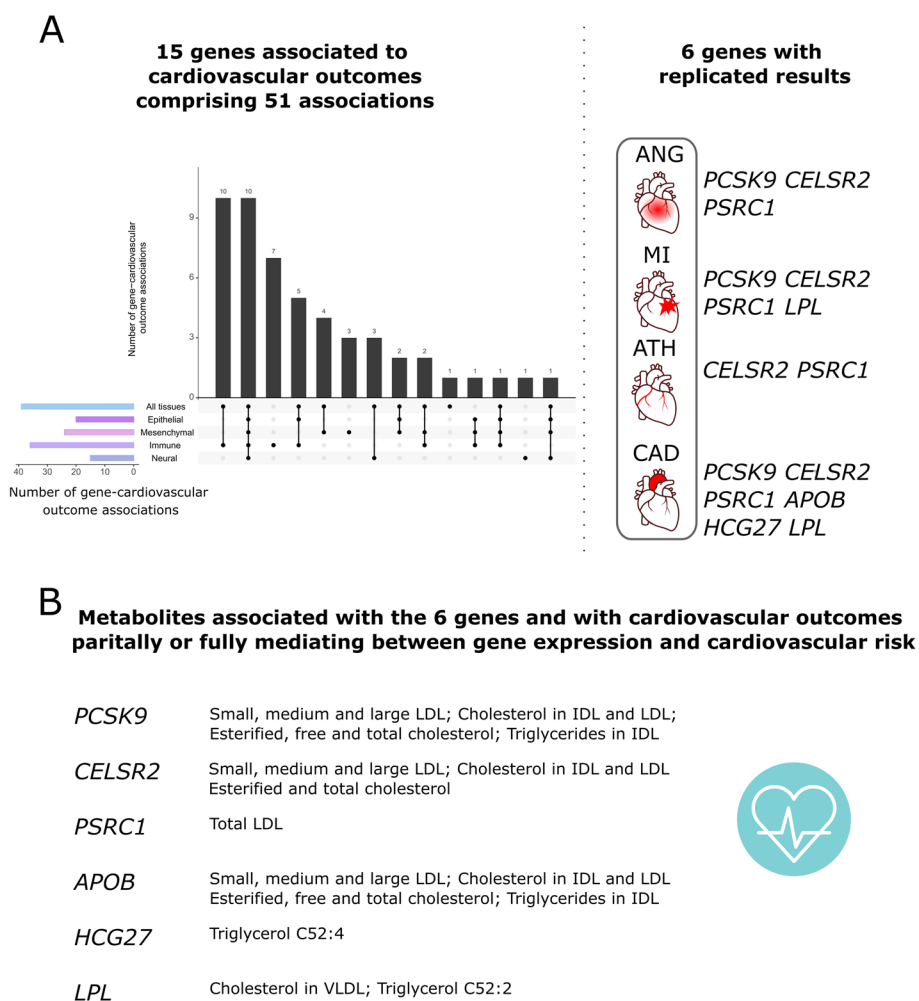
Fig. 4 Transcription-wide associated genes for cardiovascular outcomes. **A** Upset plot of the 15 gene-cardiovascular phenotypes associations per tissue category and summary of the 6 replicated genes. **B** Summary of the genetically informed mediation inference analysis among the 6 genes, metabolite levels and cardiovascular risk

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 17 of 22

of interest, that were associated with CAD risk under a MR approach using the SNPs identified in this study as instruments [57]. The metabolites associated with CAD risk were mainly atherogenic metabolites, including the steroid parameters (esterified, free and total cholesterol), levels of small, medium, large and total LDL particles, cholesterol in IDL, LDL and VLDL particles, triglycerides in IDL and levels of triglycerol C52:2 and C52:4 (Additional file 2: Table S10). Second, for each gene, we identified the main eQTL in a relevant tissue (blood for *PCSK9, CELSR2, PSRC1, HCG27* and *LPL*, and adipose subcutaneous tissue for *APOB*) of GTEx Consortium [46]. Then, we estimated the MR effect of gene expression on the metabolite levels identified in the previous step and on the CAD risk in the three datasets (CGC4D, UK Biobank and FinnGen study) [57]. As expected, all results were significant, except for the genes *APOB* and *HCG27* on the FinnGen CAD risk (Additional file 2: Table S11). Finally, we decomposed the total effects of gene expression on CAD risk into mediated effects and non-mediated effects by metabolites [62]. We observed that for the genes *APOB*, *HCG27* and *LPL*, the effects mediated by metabolites were significant and explaining the total effects, letting null contribution to non-mediated effects of gene expression of these genes on CAD risk (Fig. 4B, Additional file 2: Table S12). For the genes *PCSK9* and *CELSR2,* the effects mediated by metabolites were significant and partially explaining the total effects, thus, some effect non-mediated by the tested metabolites remained (Fig. 4B, Additional file 2: Table S12). Finally, the gene *PSRC1* showed opposite effects of mediated and non-mediated effects resulting in a total negative contribution of gene expression on CAD risk, i.e. being protective. This was the result of an increasing risk effect mediated by concentrations of total LDL, i.e. expression of *PSRC1* increases concentrations of total LDL and CAD risk, but a protective effect non-mediated by the tested metabolite, i.e. expression of *PSRC1* decreases CAD risk by independent pathways than LDL levels (Fig. 4B, Additional file 2: Table S12).

Discussion

Over the past decade, numerous GWAS have investigated metabolite levels and cardiovascular outcomes, and provided hundreds of genetic regions associated with these traits [64]. However, the interpretation of GWAS findings, the identification of the causal gene underlying susceptibility and the evaluation of relationships between genetic regions, metabolite levels and cardiovascular risk have remained challenging. In this study, to link metabolite-mediated cardiovascular genetic risk loci to key cell types and tissues, we performed integrative analyses spanning multiple GWAS on metabolite levels and cardiovascular diseases, and gene expression reference panels with different broad transcriptional programs under a genetically informed inference framework. We performed genome-wide and transcription-wide association analyses on levels of 190 plasma metabolites and metabolite ratios, from large European datasets including novel 4974 individuals of the Spanish GCAT cohort, and on risk for cardiovascular diseases, from three large European studies. In addition, because the interpretation of the TWAS results is not straightforward by the potential identification of bystander genes, we infer mediation in the relationship between gene expression of six top associated genes (*PCSK9, CELSR2, PSRC1, APOB, HCG27* and *LPL*), metabolite levels and cardiovascular risk. Interestingly, our analysis revealed that the following atherogenic metabolites were the mediators of cardiovascular risk: steroid parameters (esterified, free and total cholesterol), levels of small, medium, large and total LDL particles, cholesterol in IDL, LDL and VLDL particles, triglycerides in IDL and levels of triglycerol C52:2 and C52:4.

Some of the observed genetic associations are predicted to have functional impact, such as the rare non-synonymous SNP in *PCSK9* gene [52], the common non-synonymous SNPs in *APOB* [65], and the non-frameshifting deletion in APOB gene [66], inducing LDL cholesterol and consequently affecting CAD risk. However, the SNPs identified in *APOB* in our analysis were different from the ones observed in a Japanese study [65], pointing to population-specific functional associations.

Results from the multiomic integration approach identified 53 fine-mapped genes, most of them associated with a single metabolite class; while specificity for tissue type category was not widely observed. In contrast, in the analyses for cardiovascular outcomes, we identified 15 genes, out of the 53 genes, replicating the results for six of them, which gene expression was associated with cardiovascular outcomes showing stronger association in immune tissues. The identification of *HCG27* gene as the causal gene for CAD risk in 6p21 locus is a new finding of this study. *PCSK9* on 1p32 and *CELSR2* and *PSRC1* on 1q13 were already associated with cardiovascular risk in previous GWAS [52, 67–70]. As well, *APOB* (2p24) and *LPL* (8p21) genes have been previously associated with LDL cholesterol and triglycerides metabolism, and CAD risk [52, 65, 71, 72].

Regarding the MHC region, previous studies did not identify a clear target gene for the observed cardiovascular association [70]. HCG27 is a long noncoding RNA (lncRNA) whose locus was associated with monocyte/leukocyte ratio [73], and circulating HCG27 lncRNA has been associated with acute ischemic stroke and intracranial haemorrhage via inflammation-related signalling

pathway [74, 75]. Inflammation in tissues can disrupt normal lipid storage and mobilization processes, which would explain the previous associations between MHC region and circulating triglycerides levels [76, 77]. Our results showed that HCG27 lncRNA expression in immune cells is associated with cardiovascular risk mediated by triglycerol levels. Therefore, *HCG27* is a functional candidate gene to explain the GWAS association with CAD risk.

In the 1p32 region, the pointed gene which gene expression would be related to CAD risk was *PCSK9* in tissues enriched with immune cells, which included the liver tissue. As observed in the Human Protein Atlas browser (https://www.proteinatlas.org/), the protein encoded by *PCSK9* is mainly enriched in the liver, but also expressed in intestine, kidney and blood vessels. Our results indicate that the contribution of *PCSK9* on CAD risk is partly mediated by levels of subspecies of cholesterol and LDL particles, and partly mediated by independent pathways. Interestingly, the role of *PCSK9* reducing LDL cholesterol was found to act via endosome/lysosome LDLR degradation [78]. However, it was observed that *PCSK9* inhibitors reduced LDL cholesterol in more than 60% of initial levels on top of statin therapy, while the reduction in cardiovascular risk was about 15%, much less than expected [79]. This has been partially explained by potential less benefit in statin-treated patients, and by the lack of anti-inflammatory effect of *PCSK9* inhibitors. In addition, an increasing number of studies suggested that *PCSK9* also influences the haemostatic system, by altering platelet function and the coagulation cascade [80], and inflammatory, apoptotic and immune pathways [81]. Our results add evidence on this line where the relation between *PCSK9* and CAD risk appears to transcend lipid levels.

In the case of the 1p13 region, TWAS results identified both *CELSR2* and *PSRC1* as genes whose expression would modify lipid levels and CAD risk. Genetic studies early identified this locus as involved in lipid metabolism and CAD risk [82, 83]. Interestingly, *PSRC1* genetic variants showed significant variation among different populations, being both protective in East Asian populations but risk factors in European populations [84]. Beyond lipid metabolism, the role of *PSRC1* gene in cell growth and formation and progression of atherosclerotic plaque is well established [67]. Also, integrative functional analyses identified that *PSRC1* variants may regulate gene expression of *PSRC1, CELSR2, SORT1* genes in the liver, which alter LDL cholesterol levels and CAD risk, but also circulating granulin [85]. In particular, that study observed how the eQTL used for *PSRC1* in our study was inversely associated with protein levels of PLA2G12B, C1QTNF1, CA10 and GRN, which, in turn, can affect CAD risk. These potential mechanisms of action would explain our observations on the opposite effects of lipid mediated and non-mediated effects of *PSRC1* gene expression on CAD risk, as well the variation among different populations or type of patients, being of risk when altering levels cholesterol particles but protective when acting on independent pathways. Further functional analyses are required to determine the multifaceted contributions of *PSRC1* gene expression on CAD risk.

An interesting observation of this study is that gene expression associated with cardiovascular risk phenotypes often showed a stronger association in tissues with a high percentage of immune and mesenchymal cells. This is in accordance with atherosclerosis as a disease associated with immune-related pathways, where macrophages, monocytes and T cells play a central role in the inflammatory response, plaque development and thrombosis [86–88]. In addition, there is a growing evidence on the role of mesenchymal cells, beyond tissue structure, on epithelial cell differentiation or intestinal immune cell education [89], reinforcing the increasing evidences for the gut-cardiac paradigm in cardiovascular health [90]. Increasing evidence of immune cells action is reported. A higher cancer risk has been associated with a high immune cell tissue content [91]. Here, the gene expression profiles acting in immune cells and tissues with high content of immune and mesenchymal cells provided by this study will contribute to understanding the molecular processes driving cardiovascular risk. Gene expression of these genes can be an attractive therapeutic target to modulate atheroma progression, regression or stabilization.

Another interesting finding of this study is the enrichment of the fine-mapped genes in genes associated with nutritional and metabolic diseases. Accordingly, triglyceride-rich lipoproteins particles, LDL particles, tricarboxylic acids and branched-chain amino acids have been associated with obesity, beta-cell function, insulin resistance and incident type 2 diabetes (T2D) [92–95]. In this study, a considerably proportion of fine-mapped genes were associated with carboxylic acids, and none of these were associated with cardiovascular outcomes. Therefore, the identified carboxylic acid genes could be partially driving this enrichment.

Regarding limitations and future analyses, it should be noted that these results are restricted to European populations and cannot be generalized to other continental populations. For instance, non-synonymous SNPs in *APOB* gene identified in this study differ from SNPs identified in a Japanese study. LD patterns of European samples or co-regulating genes can originate false positives, confounded results, and complicated the interpretation of the results. Thus, transcription instrumental-based results from different ancestry populations are needed to

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 19 of 22

tease out causal genes. In addition, our results are based on the prediction of gene expression values in a TWAS framework. The predictive performance of gene expression prediction models is limited and depends on the sample sizes of the eQTL datasets used for training the statistical models. Therefore, future trials and functional analyses are needed to definitely validate the causative role of gene expression of the identified genes in the context of the proposed key cell types and tissues for metabolite levels and cardiovascular risk. Finally, since plasma was not always obtained in fasted conditions, genome-wide analyses were adjusted correcting by chylomicrons as a closely related measure of the post-prandial phase. However, this correction could obscure the genetic variants involved in chylomicrons biology [96–98].

Conclusions

We provided evidence for 53 candidate genes whose expression is associated with metabolite levels. Out of these 53 genes, the gene expression of 6 was robustly associated with cardiovascular risk phenotypes showing stronger association in immune cells and tissues with high immune component, including the novel gene regulatory target *HCG27*. In addition, we estimate the extent to which gene expression affect cardiovascular risk mediated and non-mediated by metabolite levels. These results suggest novel potential gene-drug targets to reduce the risk for cardiovascular outcomes by modifying atherogenic blood metabolite levels.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01397-2.

Additional file 1: This file contains: Workflow diagram of the performed analyses; GCAT Metabolome characterization—Analytical Methods (supplementary methodology on GCAT sample processing and metabolite profiling); Publications used to generate the additional GWAS summary statistics for metabolites (supplementary methodology on metabolite profiling of additional metabolite datasets); Cohorts used to generate the additional GWAS summary statistics for metabolites (supplementary methodology on sampling and genotyping details of the original cohorts of additional metabolite datasets); Cohorts from GWAS summary statistics for cardiovascular risk (supplementary methodology on sampling and genotyping details of the original cohorts of cardiovascular datasets); List of plots for each metabolite, which include the a) Distribution plot of metabolite levels among the 4,974 GCAT individuals, b) QQ plot of GWAS meta-analyses, c) QQ plot of TWAS overall multi-tissue results, d) Manhattan plot of GWAS meta-analyses, e) Manhattan plot of TWAS overall multi-tissue results with significant fine-mapped genes (in green), and f) LocusZoom plots for the top locus-metabolite associations (350 in total).

Additional file 2: Table S1- Characteristics of the individuals from the different studies included in the metabolite meta-analysis; Table S2- Metabolite name, platform used for metabolite profiling, group of the metabolites, number of individuals and studies meta-analysed with the GCAT cohort; Table S3- Tissue categories of tissues used to elaborate gene expression predictive models in GTEx and CEDAR datasets Table S4- Summary of the cardiovascular genome-wide association studies included in the analyses; Table S5- Independent SNP-metabolite significant associations for all

---

metabolites GWAS meta-analyses; Table S6- Genetic correlations among metabolites with significant ($P < 0.05$) SNP-based heritability estimated with the LD-Score regression approach; Table S7- Fine-mapped TWAS gene-metabolite associations of genes clustering in GWAS loci; Table S8— Gene set enrichment of metabolite's 53 fine-mapped genes for disease categories; Table S9- Significant gene-cardiovascular associations for genes fine-mapped in metabolite's TWAS analyses. Table S10- Mendelian randomization (MR) results on CAD risk for metabolite levels related to the 6 genes of interest. Table S11- Mendelian randomization (MR) results for the gene expression of the 6 genes of interest on CAD risk and on metabolite levels. Table S12- Mendelian randomization (MR) results for gene expression of the 6 genes of interest on CAD risk mediated (Med) and non-mediated (Non-Med) by metabolite levels.

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 20 of 22

Download.html), Shin et al. (https://metabolomips.org/gwas/index.php?task=download), Rhee et al. (https://www.cell.com/cell-metabolism/fulltext/S1550-4131(13)00257-X) and Draisma et al. (https://doi.org/https://doi.org/10.34894/JFWWS4). Detailed data can be found in Additional file 2: Table S1. Accession URL for GWAS summary statistics for cardiovascular outcomes are Nikpay et al. (http://www.cardiogramplusc4d.org/data-downloads/), Van der Harst et al. (https://data.mendeley.com/datasets/2zdd), Jiang et al. (https://yanglab.westlake.edu.cn/resources/u) and FinnGen (https://finngen.gitbook.io/documentation/da). Detailed data can be found in Additional file 2: Table S4. Computer code is available at the GCAT project GitHub repository (https://github.com/gcatbiobank/metabolomics_gwas) [101].

Declarations

Ethics approval and consent to participate
All research was carried out in accordance with relevant national and European guidelines and regulations. The GCAT study was carried out using anonymized data provided by the Catalan Agency for Quality and Health Assessment, within the framework of the PADRIS Program. All participants provided informed consent to participate in the study and ethical approval was obtained at the Ethical Committee at the Hospital Universitari Germans Trias i Pujol (CEI no. PI-13–020). The research conformed to the principles of the Helsinki Declaration. Additionally, detailed information regarding consents, questionnaire contents and available data can be found at www.genomesforlife.com.

Competing interests
The authors declare no competing interests.

Author details
[1]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), 17190 Salt, Girona, Spain. [2]ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908 Barcelona, Spain. [3]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute for Science and Technology, Barcelona, Spain. [4]Genomes for Life-GCAT Lab, CORE Program. Germans Trias I Pujol Research Institute (IGTP), Badalona, Spain. [5]Grup de Recerca en Impacte de Les Malalties Cròniques I Les Seves Trajectòries (GRIMTra) (IGTP), Badalona, Spain. [6]Unit of Biomarkers and Susceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), L'Hospitalet del Llobregat, 08908 Barcelona, Spain. [7]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain. [8]Department of Clinical Sciences, University of Barcelona, Barcelona, Spain.

References
1. Pirillo A, Casula M, Olmastroni E, Norata GD, Catapano AL. Global epidemiology of dyslipidaemias. Nat Rev Cardiol Springer, US. 2021;18:689–700. https://doi.org/10.1038/s41569-021-00541-4. Available from:
2. Olkowicz M, Cichon IC, Szupryczynska N, Kostogrys RB, Kochan Z, Debski J, et al. Multi-omic signatures of atherogenic dyslipidaemia: pre-clinical target identification and validation in humans. J Transl Med. 2021;19:1–23. https://doi.org/10.1186/s12967-020-02663-8. BioMed Central.
3. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. Nat Med. 2011;17(11):1410–22. https://doi.org/10.1038/nm.2538.
4. Kaptoge S, Pennells L, De Bacquer D, Cooney MT, Kavousi M, Stevens G, et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Heal. 2019;7:e1332–45.
5. Kastenmüller G, Raffler J, Gieger C, Suhre K. Genetics of human metabolism: an update. Hum Mol Genet. 2015;24:93–101.
6. Beuchel C, Becker S, Dittrich J, Kirsten H, Toenjes A, Stumvoll M, et al. Clinical and lifestyle related factors influencing whole blood metabolite levels – A comparative analysis of three large cohorts. Mol Metab. 2019;29:76–85.
7. Nassan FL, Kelly RS, Koutrakis P, Vokonas PS, Lasky-Su JA, Schwartz JD. Metabolomic signatures of the short-term exposure to air pollution and temperature. Environ Res. 2021;201:111553 Environmental Health.
8. Rhee EP, Ho JE, Chen M, Shen D, Larson MG, Ghorbani A, et al. A Genome-Wide Association Study of the Human Metabolome in a Community-Based Cohort. Cell Metab. 2014;18:130–43.
9. Shin S, Fauman EB, Petersen A, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46.
10. Draisma HHM, Pool R, Kobl M, Jansen R, Petersen A, Vaarhorst AAM, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun. 2015;6:7208.
11. Kettunen J, Demirkan A, Würtz P, Draisma HHM, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7: 11122.
12. Long T, Hicks M, Yu H, Biggs WH, Kirkness EF, Menni C, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet Nat Pub Gr. 2017;49:568–78.
13. Gallois A, Mefford J, Ko A, Vaysse A, Julienne H, Ala-korpela M, et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat Commun. 2019;10:4788. https://doi.org/10.1038/s41467-019-12703-7. Springer, US.
14. Margoliash J, Fuchs S, Li Y, Zhang X, Massarat A, Goren A, et al. Polymorphic short tandem repeats make widespread contributions to blood and serum traits. Cell Genomics. 2023;3:100458. https://doi.org/10.1016/j.xgen.2023.100458. The Author(s).
15. Cadby G, Giles C, Melton PE, Huynh K, Mellett NA, Duong T, et al. Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. Nat Commun. 2022;13:1–17.
16. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet. 2020;11: 424.
17. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Publ Gr. 2016;48:245–52. https://doi.org/10.1038/ng.3506. Available from.
18. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. Plos Genet. 2019;15:e1007889. https://doi.org/10.1371/journal.pgen.1007889.
19. Breschi A, Muñoz-Aguirre M, Wucher V, Davis CA, Garrido-Martín D, Djebali S, et al. A limited set of transcriptional programs define major cell types. Genome Res. 2020;30:1047–59.
20. Li B, Veturi Y, Bradford Y, Verma SS, Verma A, Lucas AM, et al. Influence of tissue context on gene prioritization for predicted transcriptome-wide association studies 1. Introduction Improving antiretroviral therapy (ART) efficacy and safety is an ongoing goal for addressing the HIV pandemic. According to the Joi. Pac Symp Biocomput. 2019;24:296–307.
21. Li L, Chen Z, von Scheidt M, Li S, Steiner A, Güldener U, et al. Transcriptome-wide association study of coronary artery disease identifies novel susceptibility genes. Basic Res Cardiol. 2022;117:1–20. https://doi.org/10.1007/s00395-022-00917-8. Springer, Berlin Heidelberg.
22. Highland HM, Wojcik GL, Graff M, Nishimura KK, Hodonsky CJ, Baldassari AR, et al. Predicted gene expression in ancestrally diverse populations leads to discovery of susceptibility loci for lifestyle and cardiometabolic traits. Am J Hum Genet. 2022:1–11. https://doi.org/10.1016/j.ajhg.2022.02.013. American Society of Human Genetics.
23. Thompson M, Gordon MG, Lu A, Tandon A, Halperin E, Gusev A, et al. Multi-context genetic modeling of transcriptional regulation resolves novel disease loci. Nat Commun Springer, US. 2022;13:1–15.
24. Zhao Q, Liu R, Chen H, Yan X, Dong J, Bai M, et al. Transcriptome-wide association genes for coronary atherosclerosis. Front Cardiovasc Med. 2023:1–8. https://doi.org/10.3389/fcvm.2023.1149113.
25. Ndungu A, Payne A, Torres JM, van de Bunt M, McCarthy MI. A Multi-tissue Transcriptome Analysis of Human Metabolites Guides Interpretability of Associations Based on Multi-SNP Models for Gene Expression.

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 21 of 22

Am J Hum Genet ElsevierCompany. 2020;106:188–201. https://doi.org/10.1016/j.ajhg.2020.01.003. Available from:

26. de Leeuw C, Werme J, Savage JE, Peyrot WJ, Posthuma D. On the interpretation of transcriptome-wide association studies. PLoS Genet. 2023;19:1–23. Available from: https://doi.org/10.1371/journal.pgen.1010921.

27. Obón-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galván-femenía I, et al. GCAT | Genomes for life: a prospective cohort study of the genomes of Catalonia. BMJ Open. 2018;8: e018324.

28. Galván-Femenía I, Obón-Santacana M, Piñeyro D, Guindo-Martinez M, Duran X, Carreras A, et al. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. J Med Genet. 2018;55:765–78.

29. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9:179–81.

30. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5:e1000529. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2689936&tool=pmcentrez&rendertype=abstract. Cited 2014 Jan 22.

31. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68–74.

32. Huang J, Howie B, Mccarthy S, Memari Y, Walter K, Min JL, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat Commun Nature Publishing Group. 2015;6:8111.

33. Deelen P, Menelaou A, Van Leeuwen EM, Kanterakis A, Van Dijk F, Medina-gomez C, et al. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands.' Eur J Hum Genet. 2014;22:1321–6.

34. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.

35. Guindo-Martínez M, Amela R, Bonàs-Guarch S, Salvoro C, Miguel-Escalada I, Carey CE, et al. The impact of non-additive genetic associations on age-related complex diseases. Nat Commun Springer, US. 2021;12:2436. https://doi.org/10.1038/s41467-021-21952-4. Available from:

36. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res. 2018;46:D608-17.

37. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

38. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2922887&tool=pmcentrez&rendertype=abstract. Cited 2013 May 25.

39. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481–7.

40. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011. The American Society of Human Genetics.

41. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012;28:2540–2.

42. Urbut SM, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet Springer, US. 2019;51:187–95. https://doi.org/10.1038/s41588-018-0268-8. Available from:

43. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun Springer, US. 2017;8:1826. https://doi.org/10.1038/s41467-017-01261-5. . Available from:

44. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–7.

45. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9:1825.

46. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (80-). 2020;369:1318–30.

47. Barbeira AN, Melia OJ, Im HK, Wheeler HE, Bonazzola R, Wang G, et al. Fine-mapping and QTL tissue-sharing information improves the reliability of causal gene identification. Genet Epidemiol. 2020;44:854–67.

48. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charloteaux B, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun. 2018;9:2427.

49. Díez-Obrero V, Dampier CH, Moratalla-Navarro F, Devall M, Plummer SJ, Díez-Villanueva A, et al. Genetic Effects on Transcriptome Profiles in Colon Epithelium Provide Functional Insights for Genetic Risk Loci. Cell Mol Gastroenterol Hepatol. Elsevier Inc; 2021;12:181–97. Available from: https://doi.org/10.1016/j.jcmgh.2021.02.003.

50. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, et al. Probabilistic fine-mapping of transcriptome-wide association studies. Nat Genet. 2019;51:675–82. https://doi.org/10.1038/s41588-019-0367-1. Springer, US.

51. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48:D845–55.

52. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47:1121–30.

53. FinnGen. FinnGen documentation of R6 release. 2022. Available from: https://finngen.gitbook.io/documentation/. Cited 2022 Jan 15.

54. van der Harst P. CAD GWAS in UK Biobank [Internet]. Mendeley Dataset. 2017. p. V1. Available from: https://data.mendeley.com/datasets/2zdd47c94h/1. Cited 2021 Jan 15.

55. Jiang L, Zheng Z, Fang H, Yang J. A generalized linear mixed model association tool for biobank-scale data. Nat Genet. 2021;53:1616–21 Springer, US.

56. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23:1–10. Available from: http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddu328.

57. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–65. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4377079andtool=pmcentrezandrendertype=abstract.

58. Burgess S, Thompson SG. Avoiding bias from weak instruments in mendelian randomization studies. Int J Epidemiol. 2011;40:755–64.

59. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40:597–608.

60. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;512–25. Available from: http://www.ije.oxfordjournals.org/cgi/doi/10.1093/ije/dyv080.

61. Bowden J, Smith GD, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.

62. Burgess S, Daniel RM, Butterworth AS, Thompson SG. Network Mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol. 2015;44:484–95.

63. Hagenbeek FA, Pool R, van Dongen J, Draisma HHM, Jan Hottenga J, Willemsen G, et al. Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. Nat Commun. 2020;11:39.

64. Kessler T, Schunkert H. Coronary artery disease genetics enlightened by genome-wide association studies. JACC Basic to Transl Sci. 2021;6:610–23. https://doi.org/10.1016/j.jacbts.2021.04.001. Elsevier.

65. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nat

Carreras-Torres *et al. Genome Medicine*   (2024) 16:122

Page 22 of 22

Genet Springer, US. 2020;52:1169–77. https://doi.org/10.1038/s41588-020-0705-3. Available from:

66. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature. 2021;600:675–9.

67. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide Association Analysis of Coronary Artery Disease. N Engl J Med. 2007;357:443–53.

68. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009;41:334–41. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2681011&tool=pmcentrez&rendertype=abstract. Cited 2013 May 25.

69. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011;43:333–40.

70. Davies RW, Wells GA, Stewart AFR, Erdmann J, Shah SH, Ferguson JF, et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. Circ Cardiovasc Genet. 2012;5:217–25.

71. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transen-dothelial migration pathways to coronary artery disease. Nat Genet. 2017;49:1392–7 Nature Publishing Group.

72. Van Der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ Res. 2018;122:433–43.

73. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell. 2020;182:1214-1231.e11.

74. Huang L, Li X, Chen Z, Liu Y, Zhang X. Identification of inflammation-associated circulating long non-coding RNAs and genes in intracranial aneurysm rupture-induced subarachnoid hemorrhage. Mol Med Rep. 2020;22:4541–50.

75. Zhang L, Liu B, Han J, Wang T, Han L. Competing endogenous RNA network analysis for screening inflammation-related long non-coding RNAs for acute ischemic stroke. Mol Med Rep. 2020;22:3081–94.

76. Teslovich TM, Musunuru K, Smith AV, Edmondson CA, Stylianou IM, et al. Biological, clinical, and population relevance of 95 Loci for blood lipids. Nature. 2010;466:707–13.

77. Willer CJ, Sanna S, Jackson AU, Scuteri A, Lori L, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet. 2008;40:161–9.

78. Jang HD, Lee SE, Yang J, Lee HC, Shin D, Lee H, et al. Cyclase-associated protein 1 is a binding partner of proprotein convertase subtilisin/kexin type-9 and is required for the degradation of low-density lipoprotein receptors by proprotein convertase subtilisin/kexin type-9. Eur Heart J. 2020;41:239–52.

79. Waters DD, Hsue PY. PCSK9 inhibition to reduce cardiovascular risk. Circ Res. 2017;120:1537–9.

80. Chong S, Mu G, Cen X, Xiang Q, Cui Y. Effects of PCSK9 on thrombosis and haemostasis in a variety of metabolic states: Lipids and beyond (Review). Int J Mol Med. 2024;53:57.

81. Tang Y, Li SL, Hu JH, Sun KJ, Liu LL, Xu DY. Research progress on alternative non-classical mechanisms of PCSK9 in atherosclerosis in patients with and without diabetes. Cardiovasc Diabetol. 2020;19:33. https://doi.org/10.1186/s12933-020-01009-4. BioMed Central.

82. Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, et al. Genetic Variants Influencing Circulating Lipid Levels and Risk of Coronary Artery Disease. Arterioscler Thromb Vasc Biol. 2010;30:2264–76.

83. Battram T, Hoskins L, Hughes DA, Kettunen J, Ring SM, Smith GD, et al. Coronary artery disease, genetic risk and the metabolome in young individuals. Wellcome Open Res. 2019;3:114.

84. Hu Y, Qiu S, Cheng L. Integration of Multiple-Omics Data to Analyze the Population- Specific Differences for Coronary Artery Disease. Comput Math Methods Med. 2020;56:427.

85. Chai T, Wang Z, Yang X, Qiu Z, Chen L. PSRC1 May Affect Coronary Artery Disease Risk by Altering CELSR2, PSRC1, and SORT1 Gene Expression and Circulating Granulin and Apolipoprotein B Protein Levels. Front Cardiovasc Med. 2022;9: 763015.

86. Goikuria H, Vandenbroeck K, Alloza I. Inflammation in human carotid atheroma plaques. Cytokine Growth Factor Rev. 2018;39:62–70.

87. Barrett TJ. Macrophages in Atherosclerosis Regression. Arterioscler Thromb Vasc Biol. 2020;40:20–33.

88. Tan L, Xu Q, Shi R, Zhang G. Bioinformatics analysis reveals the landscape of immune cell infiltration and immune-related pathways participating in the progression of carotid atherosclerotic plaques. Artif Cells Nanomedicine Biotechnol. 2021;49:96–107. https://doi.org/10.1080/21691401.2021.1873798. Taylor and Francis.

89. Kurashima Y, Yamamoto D, Nelson S, Uematsu S, Ernst PB, Nakayama T, et al. Mucosal mesenchymal cells: Secondary barrier and peripheral educator for the gut immune system. Front Immunol. 2017;8: 1787.

90. Akshay A, Gasim R, Ali TE, Kumar YS, Hassan A. Unlocking the Gut-Cardiac Axis: A Paradigm Shift in Cardiovascular Health. Cureus. 2023;15: e51039.

91. Palomero L, Galván-Femenía I, de Cid R, Espín R, Barnes DR, CIMBA, et al. Immune cell associations with cancer risk. iScience. 2020;23:101296.

92. Sokooti S, Flores-Guerrero JL, Kieneker LM, Heerspink HJL, Connelly MA, Bakker SJL, et al. HDL Particle Subspecies and Their Association with Incident Type 2 Diabetes: The PREVEND Study. J Clin Endocrinol Metab. 2021;106:1761–72.

93. Mycielska ME, James EN, Parkinson EK. Metabolic alterations in cellular senescence: the role of citrate in ageing and age-related disease. Int J Mol Sci. 2022;23:3652.

94. Vanweert F, Schrauwen P, Phielix E. Role of branched-chain amino acid metabolism in the pathogenesis of obesity and type 2 diabetes-related metabolic disturbances BCAA metabolism in type 2 diabetes. Nutr Diabetes. 2022;12:35 Springer US.

95. Bloomgarden Z. Diabetes and branched-chain amino acids: What is the link? J Diabetes. 2018;10:350–2.

96. Frazier-Wood AC, Kabagambe EK, Borecki IB, Tiwari HK, Ordovas JM, Arnett DK. Preliminary evidence for an association between LRP-1 genotype and body mass index in humans. PLoS ONE. 2012;7:8–11.

97. Frazier-Wood AC, Kabagambe EK, Wojczynskid MK, Boreckid IB, Tiwarib HK, Smithe CE, et al. The Association Between LRP-1 Variants and Chylomicron Uptake After a High Fat Meal. Nutr Metab Cardiovasc Dis. 2013;23:1154–8.

98. Desmarchelier C, Martin JC, Planells R, Gastaldi M, Nowicki M, Gon-calves A, et al. The postprandial chylomicron triacylglycerol response to dietary fat in healthy male adults is significantly explained by a combination of single nucleotide polymorphisms in genes involved in triacylglycerol metabolism. J Clin Endocrinol Metab. 2014;99:484–8.

99. GCAT|Genomes for Life. GCAT study genetic deposited in the European Genome-phenome Archive (EGA) under study accession EGAS00001003018. Available from: https://ega-archive.org/studies/EGAS00001003018.

100. GCAT|Genomes for Life. GCAT Data Access. Available from: http://www.gcatbiobank.org/investigadors/en_gcat-data-access/.

101. GCAT|Genomes for Life. GCAT project GitHub repository. Available from: https://github.com/gcatbiobank/metabolomics_gwas.

Publisher's Note

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