Relevance of Multi-Omics Studies in Cardiovascular Diseases

Paola Leon-Mimila, Jessica Wang and Adriana Huertas-Vazquez*

Division of Cardiology, David Geffen School of Medicine, Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

Cardiovascular diseases are the leading cause of death around the world. Despite the larger number of genes and loci identified, the precise mechanisms by which these genes influence risk of cardiovascular disease is not well understood. Recent advances in the development and optimization of high-throughput technologies for the generation of “omics data” have provided a deeper understanding of the processes and dynamic interactions involved in human diseases. However, the integrative analysis of “omics” data is not straightforward and represents several logistic and computational challenges. In spite of these difficulties, several studies have successfully applied integrative genomics approaches for the investigation of novel mechanisms and plasma biomarkers involved in cardiovascular diseases. In this review, we summarized recent studies aimed to understand the molecular framework of these diseases using multi-omics data from mice and humans. We discuss examples of omics studies for cardiovascular diseases focused on the integration of genomics, epigenomics, transcriptomics, and proteomics. This review also describes current gaps in the study of complex diseases using systems genetics approaches as well as potential limitations and future directions of this emerging field.

Keywords: multi-omics, cardiovascular disease, heart disease, systems biology, data integration

INTRODUCTION

Coronary artery disease (CAD) is the most common cause of cardiovascular death (1). Studies conducted in twins (2, 3) and in the general population have estimated a heritability of CAD at ∼40–50% (4). In addition, genome-wide association studies (GWAS) have identified more than 150 genetic loci associated with CAD risk (5–18). Although GWAS studies have been successful on identifying common DNA variation implicated in cardiovascular diseases, they provide little or no molecular evidence of gene causality. In this context, the premise that rare genetic variation could have stronger functional effects on disease manifestation still is arguable (19). This realization has motivated researchers to integrate genetics studies with additional high-throughput data designed to interrogate the transcriptome, epigenome, proteome, metabolome, etc.
Recent studies have implemented the integration of multi-omics data to accelerate the identification of novel mechanisms for complex diseases and understand the dynamics of disease manifestation (20–23). The relevance of integrating multi-omics data and the current statistical tools available for data integration have been reviewed in detail elsewhere (24–34). In this review, we summarize the state-of-the-art of multi-omics studies conducted in mice and humans to understand the molecular mechanisms underlying cardiovascular diseases including CAD (35–47), stroke (42, 48), heart failure (13, 49, 50), cardiac hypertrophy (13, 51), aortic valve disease (52, 53), and heart regeneration (54). We also discuss the gaps of multi-omics studies including the utility of generating multi-omics data in animal models, the importance of sex stratification on gene discovery, the inclusion of diverse populations and the integration of metabolomics and metagenomics with other omics platforms. Finally, we discuss future directions of multi-omics approaches for cardiovascular diseases and their importance in the era of precision health.

**MULTI-OMICS STUDIES FOR THE INVESTIGATION OF CARDIOVASCULAR DISEASE**

The simultaneous integration of multi-omics approaches including but not limited to genomics, epigenomics, transcriptomics, proteomics, and metabolomics (Figure 1), represents a powerful approach for understanding the mechanisms connecting identified genetic variation to cardiovascular diseases with gene causality, where many sources of variability are integrated into statistical models to identify key drivers and pathways that have the largest contribution to the disease (25). Importantly, most of the risk variants associated with CAD or other cardiovascular diseases (5, 7, 14, 17, 18, 37, 55, 56) identified by GWAS are located in noncoding regions of the genome (intronic or intergenic), suggesting that these variants are likely to affect cis or trans regulatory elements that bind transcription factors, enhancers or promoters (57). Previous multi-omics studies for CAD were mainly focused on the integration of GWAS data with global transcriptomics using eQTL analysis. In recent years, high-throughput technology have further facilitated the integration of omics data for the identification of causal genes and molecular mechanisms involved in the development of cardiovascular events in mice (13, 37, 39, 41, 58) and humans (36–39, 48) (Table 1).

**SUCCESS STORIES OF MULTI-OMICS STUDIES IN CARDIOVASCULAR DISEASES**

Although there have been few studies integrating multi-omics profiles for the investigation of mechanisms associated with cardiovascular diseases, this approach has revealed the potential function of previously identified GWAS loci and respective mechanisms involved in these common diseases. In this section, we summarize recent studies using multi-omics approaches focusing on the integration of genomics, epigenomics, transcriptomics, and proteomics.

**Genomics, Transcriptomics, and Epigenomics**

There is a large body of literature linking genetic variation with gene expression and/or epigenetic marks to understand the potential mechanisms of identified DNA variants in disease manifestation. One example on the integration of genomics with transcriptomics is a study conducted to investigate the role of the 9p21 locus (63), which was identified as one of the most significant loci for CAD in previous GWAs (64, 65). The association of CAD with this locus have been consistently replicated in multiple studies (56, 66), although the causal link of this locus remained unclear. This locus contains several genes including CDKN2A (encoding cyclin p14, p16), CDKN2B (encoding cyclin p15), MTAP (encoding methylthioadenosine phosphorylase), and the long non-coding RNA ANRIL. Integration of genetic and transcriptomic data led to the identification of ANRIL as the top candidate causal gene for CAD at the 9p21 region (63). Functional studies in cell lines showed possible mechanisms that could explain the role of 9p21 in CAD (67, 68). For instance, a previous study showed that alleles at the 9p21 locus were associated with different isoforms of ANRIL (linear or circular isoforms), where linear transcripts were associated with atherosclerosis and circular transcripts were protective against atherosclerosis. This process is mediated through the expression of multiple genes regulated in both, cis and trans (69, 70). Moreover, a recent study showed that ANRIL (DQ485454) is involved in endothelial cells functions important to the development of CAD including monocyte adhesion to endothelial cells, trans-endothelial monocyte migration, and endothelial cell migration (71).

Another example is the investigation of the region of the gene cluster CELSR2-PSRC1-MYBPHL-SORT at the 1p13.3 locus associated with low-density lipoprotein cholesterol (LDL-C) levels and cardiovascular risk (55, 72, 73). Incorporation of eQTL analysis also showed that SNPs associated with a lower risk of CAD in the 1p13.3 locus were associated with an increased gene expression of SORT1, PSRC1, and CELSR2, with SORT1 displaying the largest expression change in the liver (73, 74). This finding allowed the construction of new hypothesis to elucidate the molecular mechanism of the 1p13.3 locus on CAD development. Studies of SORT1 and PSRC1 overexpression in mouse models of hyperlipidemia showed that, while PSRC1 overexpression had no metabolic effects, SORT1 overexpression led to a significant reduction in plasma LDL-C and very low-density lipoprotein (VLDL) particle levels by modulating hepatic VLDL secretion, suggesting an important role of SORT1 in CAD (74). Finally, a similar omics approach was applied to identify genes associated with isoproterenol-induced hypertrophy and heart failure in the Hybrid Mouse Diversity Panel (HMDP) (13, 22, 23, 41, 75–83). The integration of genomic information and cardiac transcriptome enabled the identification of several candidate causal genes that determined the degree of cardiac
Leon-Mimila et al. Multi-Omics and Cardiovascular Disease

**FIGURE 1** | Multi-omics approach to identify the causal gene associated with LDL-C levels and CAD risk at the 1p13 locus. (A) GWAs meta-analysis showed several SNPs at the 1p13 locus strongly associated with LDL-C levels (\(p = 1.0 \times 10^{-170}\)) and CAD risk. The 1p13 locus contains several genes (squares). The most significantly associated haplotype for LDL-C comprise six SNPs in high linkage disequilibrium (LD) and is located between CELSR1 and PSR1 genes. (B) Liver eQTL analysis showed the minor haplotype significantly associated with higher expression of CELSR1, PSR1, and SORT1 genes with SORT1 gene showed the largest difference modified from Musunuru et al. (74). (C) By using luciferase assays and ENCODE database it was identified a common polymorphism at the 1p13 locus, rs12740374 that alters the expression of the SORT1 gene in liver with the minor allele (T) creating a C/EBP (CCAAT/ enhancer binding protein) transcription factor binding site and the major allele (G) disrupting it. The C/EBP transcriptional factor regulates the expression of hepatic genes involved in metabolism. (D) Functional approaches for SORT1 using small interfering RNA (siRNA) knockdown and viral overexpression in mouse liver showed that SORT1 results in significant changes in plasma LDL-C and very low-density lipoprotein (VLDL) particle levels by modulating hepatic VLDL secretion.

hypertrophy. Specifically, **Hes1** was predicted to be involved in the progression of heart damage in cardiac hypertrophy (13). This study showed that knocking down **Hes1** in ventricular myocytes resulted in a reduction of up to 90% hypertrophy, confirming the role of **Hes1** in cardiac hypertrophy (13). More recently, several studies have demonstrated that epigenetic modifications are associated with CAD risk (38, 42, 43, 47, 49, 59, 61, 62, 84, 85), and other CVD related risk factors (61, 62, 84). Epigenetic changes that have been investigated in the context of CVD include DNA methylation (38, 43, 49), chromatin organization (42), and microRNAs (47). In recent years, efforts have been conducted to identify interactions between functional non-coding active elements of the genome and enhancers, defined as cis-acting DNA sequences that can increase the transcription of genes (60, 61, 86). Several methods have been developed for the identification of these interactions including, chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq), chromatin conformation capture (3C, HiC), and most recently, chromatin interaction paired-end tagging (ChIA-PET). These technologies offer the advantage to identify genome-wide protein-DNA interactions.

**Adding Another Layer: Proteomics**

The incorporation of protein expression profiles into the multi-omics studies for CAD has been less explored compared with multi-omics studies incorporating mRNA expression (43–45, 47, 51–54). This may be due to the costs and the highly specialized expertise required for instrument operation, data acquisition,
| References                  | Phenotypes                  | Population of study                                      | Omic strategy                              | Tissue | Analysis strategy           | Main findings                                                                                                      | Genes involved                                                                                           | Functional confirmation                                                                 |
|-----------------------------|-----------------------------|----------------------------------------------------------|--------------------------------------------|--------|-----------------------------|-------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Santolini et al. (13)       | Isoproterenol-induced cardiac hypertrophy and heart failure | Mice (HMDP) 100 genetically diverse strains of mice      | Genomics (genomic diversity)              | H      | Correlation-based method    | Identification of 36 genes associated with severity of cardiac hypertrophy                                        | Rhl, Wdr1, Nppb, Atp6v0a1, Ankrd1, E4fa1, Dr (HB-EGF), Kcnip2, Pcdhgc4, Hes1, 493080406Rik, Akap9, 2310022B05Rik, Botaf1, Tc3L, Npsap3B, Gsz, Knk23, Tsan17, Trin2, Cab39L, Pfr (Cavin-1), ventricular myocytes | Knockdown of Hes1 reduces hypertrophy by 80-90% in neonatal rat cardiac myocytes |
| Foroughi Asl et al. (36)    | CAD                         | CAD patients from the Stockholm Atherosclerosis Gene Expression (STAGE) study | Genomics (microarray platform, Affymetrix) | B, AAW, MAM, LIV, SKLM, SF, VAF | Cis- and trans-gene regulation by GWAS risk loci across tissues and CAD phenotypes | Identification of 3 master regulators of CAD across 7 tissues                                                 | FLYWCH1, PSORSIC3 and G3BP1                                                      | Knockdown of FLYWCH1, PSORSIC3, G3BP1 genes affect cholesterol-ester accumulation in foam cells |
| Braenne et al. (37)         | CAD                         | Mice (HMDP)                                              | Genomics (microarray platform, Affymetrix) | LIV, SF, and M | GWAS and eQTL analysis   | The majority of the GWAS loci for CAD affect gene expression (41%)                                            | LIPA, TOM1L2, GALNT4, SERPINH1, VAMP8, VAMP5, GGCX, PSCR1, CELSR2, SORT1, DRG2, C17orf39, MYO15A, TOM1L2, SREBF1, mir-224, hsa-miR-130a-5p, hsa-miR-4722-5p, hsa-miR-3198, hsa-miR-3197-3p, miR-378a-5p | NA |
| Zhao et al. (46)            | Carotid plaque, Stroke      | Gene-expression profiles of 11 publically gene expression datasets of carotid plaque (n = 1,546), GWA studies of ischemic stroke from the International Stroke Genetics Consortium | Genomics (microarray platform, Illumina)  | H      | Marker Set Enrichment Analysis (co-expression modules) | Seventeen co-expression modules were enriched for stroke. Enriched modules for stroke were associated with toll-like receptor pathway, homocysteine metabolism and phagosome formation and maturation | LDLR, APOE, SCARB1, NOS3, CSNK2A1, HTRA1, LRPL1, COL4A1, FN1, RELA, TNF, SHC1, LRP1, LYN, SYK, IGF1R, SHC1, IL6R, CXCR4, LCAT, LDLR, PLTR, APP, SCH1, RELA, FN1, TNF, FN1, PCSK9, TRI1B3, CXCR4, and CCR1 | NA |
| Lempainen et al. (48)       | CAD                         | GWAS studies and exome array studies for CAD. eQTL STAGE study | Genomics (microarray platform, Illumina)  | B, AAW, SKLM, SF, VAF | Construction of network modules for tissue-specific gene-protein interactions affected by genetic variance in CAD risk loci | Identification of modules with tissue-specific activity associated with CAD. Most of the modules were druggable. The top modules were implicated in extracellular matrix organization and disassembly, blood coagulation, or platelet degranulation/activation process | NA |

(Continued)
TABLE 1 | Continued

| References | Phenotypes | Population of study | Omic strategy | Tissue | Analysis strategy | Main findings | Genes involved | Functional confirmation |
|------------|------------|---------------------|---------------|--------|------------------|---------------|----------------|----------------------|
| Franzen et al. (38) | CAD | Patients with CAD from the STARNET study Road Epigenomics Consortium | Genomics (microarray platform, Illumina) | B, MAM, AOR, SF, VAF, SKLM, LIV | Cis- and trans-gene regulation across different tissues and CAD phenotypes | Tissue-specific gene-regulatory effects of CAD-associated SNPs identified by GWAS. Identification of 26 key drivers regulated in cis-trans by CAD SNPs | FAM117B, LiPA, SARS, ATP5G1, GGCX, CARF, ICA1L, SH2B3, AC023271.1, RPL7P14, MAT2A, EDNRA, LINCO0310, SLC22A5, N75C2, FES, USP89, ADAMTS7, FURIN, PSMAS, ABCG5, CNM32, SLCSA3, CA1FD, ZNF76, TCF21, PSRC1, and PDGFD | NA |
| Liu et al. (59) | CAD | HCASMCs from 52 unrelated donors. | Genomics (microarray platform, Illumina) | HCASMCs | Jointly eQTL modeling and GWAS analyses | Identification of 5 genes that modulate CAD risk via HCASMCs. | SIPA1, TCF21, SAD3D, FES, PDGFR | NA |
| Haitjema et al. (42) | CAD, Stroke | GWAS of METASTROKE and CARDioGRAMplusC4D | Genomics (microarray platform, Illumina) | M, CEC | Association of eQTLs with chromatin interaction | Integrative analysis of gene expression and chromatin conformation to elucidate mechanisms involved in atherosclerosis | MA3, PSRC1, SORT1, GGCX, VAMP5, VAMP8, NBEAL1, WDR12, MRAS, PHACTR1, TRB1, CDKN2A, CDKN2B, KIAA1462, LIPA, COL4A1, COL4A2, PEMT, RASD1, SMG6, UBE2Z, LDLR | NA |
| Lee et al. (60) | Heart failure | 135 patients with dilated cardiomyopathy | Transcriptionomics (mRNA sequencing, Illumina) | H | Methylation-expression quantitative trait locus analysis | Integration of methylation and gene expression data identified enrichment of cell adhesion, cardiac development, and muscle function in HF | PLXNA2, RGS3, NPPA, NPPB, B9D1, + | doublecortin-like kinase 2 and neurotrimin |
| Meder et al. (49) | Hypertension MI Stroke Thrombosis Aorthytymia | 729 subjects from the Northern Sweden Population Health Study | Epigenomics (microarray platform, Illumina) | B | Integration of EWAS and ChiA-PET data | Identification of 196 genes associated with cardiac-related traits | ESRRG, ST6GALNAC5, RYR2, NNNAT2, EphA2, TGFB2, ABCG5, FMRNL2, Dysf, Myb1, Mecom, WNT7A, Sox2, Hand2, F2RL1, KcnN2, ME1* CPT1A and SREBF1 (TGs) HCR24 (LDL-C) ABCG1 (HDL-C) | NA |
| Dekkers K et al. (62) | Blood lipids | 3,296 subjects from the Biobank Based Integrative Omics Study | Transcriptionomics (mRNA sequencing, Illumina) | B | Integration of EWAS and gene expression | Identification of CpGs associated with the expression of lipids | ATP1B1, NME7, Camsap2, Ddx59, Lm001, Tnn1, Tnks1, Tbx5at1, SerpinH1, Scarb1, Trf4 Hr, Pecam1, Procr | NA |
| Howson JM, et al. (43) | CAD | 88,192 CAD cases 162,544 controls including CARDioGRAMplusC4D database | Genomics (microarray platform, Illumina, Affymetrix) | 30 cells/tissues including P, B, LIV, SF, VAF, H, and DT | Genomic meta-analysis, eQTL, pQTL. Enrichment analysis (Ingenuity Pathway Analysis Software) | Integrative analysis showed enrichment of genes involved in biological processes active in the arterial wall as cellular adhesion, leukocyte migration, vascular smooth muscle cell differentiation, coagulation, inflammation, and atherosclerosis | ATPIB1, NME7, CAMSAP2, DDX59, LMO01, TN1S1, TBX5AS1, SERPINH1, SCARB1, TRIP4 HP, PECAM1, PROCR | NA |
| References     | Phenotypes               | Population of study                                                                 | Omic strategy                                                                 | Tissue | Analysis strategy                          | Main findings                                                                 | Genes involved                                                                 | Functional confirmation |
|----------------|--------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|--------|--------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------|------------------------|
| Yao C, et al.  | CAD                      | 6,861 subjects from the Framingham Heart Study and CARDioGRAMplusC4D               | Genomics (microarray platform, Illumina, Affymetrix)                          | P      | Multi-stage strategy of proteomic analysis| pQTL analysis identified six causal proteins for CHD                         | LPA, BCHE, PON1, MCAM, MPO, Cystatin C                                        | NA                     |
|                |                          |                                                                                     | Transcriptomics (microarray platform, Affymetrix)                             |        |                                            |                                                                                |                                                                                |                        |
|                |                          |                                                                                     | Proteomics (multiplexed aptamer based alinity proteomics platform, Luminex)   |        |                                            |                                                                                |                                                                                |                        |
| Chen G, et al. | CAD, MI                  | 7,242 participants from the Framingham Heart Study                                  | Genomics (microarray platform, Illumina, Affymetrix)                          | P      | Cis- and trans-protein regulation by GWAS  | Identification of 210 pQTLs for 12 proteins associated with CAD and MI        | CELSR2/SORT1 locus (granulin)                                                  | NA                     |
|                |                          |                                                                                     | Targeted proteomics (bead-based multiplex immunoassays, Luminex)             |        |                                            |                                                                                |                                                                                |                        |
| Fernandez, M, | CAD                      | Public databases of human samples                                                   | Genomics (microarray platform, Illumina, Affymetrix)                          | ART, B, H, and LIV | Supervised development of a multi-omics integrative molecular model | Integrative analysis of omics studies showed enrichment of lipid metabolism, extracellular matrix remodeling, inflammation, and cardiac hypertrophy pathways | LCAT, FABP1, FASN, APOA1, FASN, mir-1303 (PPARA and APOA1), mir-1303 (FASN)  | NA                     |
| et al. (47)    |                          |                                                                                     | Transcriptomics (microarray platform, Illumina)                              |        |                                            |                                                                                |                                                                                |                        |
| Lau E, et al.  | Cardiac hypertrophy      | Mice (inbred from six diverse genetic backgrounds)                                 | Transcriptomics (microarray platform, Illumina)                              | H      | Clustering of co-expression               | Modules associated with heart hypertrophy across the mouse strains were involved in biological processes including cell adhesion, glycolytic process, actin filament organization, translation, and sodium ion transport | ANXA2, ANXA5, COL4A2, LDHA, and PGAM1                                        | NA                     |
|                |                          |                                                                                     | Proteomics (LC-MS/MS platform, Thermo)                                         |        |                                            |                                                                                |                                                                                |                        |
|                |                          |                                                                                     | Proteome dynamics                                                             |        |                                            |                                                                                |                                                                                |                        |
| Schlotter F, et al. (52) | Calcific aortic valve disease | 25 human stenotic aortic valves                                                      | Transcriptomics (mRNA sequencing, Illumina)                                  | AV     | Correlation of gene and protein expression differentiated between calcification stage. Protein-protein interaction | Identification of novel regulatory networks for CAVD                           | SOD3, MGR SERPINA1, WWF, C8A, C8B, SLPI, ELANE, HLA-DRA, and CD14            | NA                     |

(Continued)
| References | Phenotypes | Population of study | Omic strategy | Tissue | Analysis strategy | Main findings | Genes involved | Functional confirmation |
|------------|------------|---------------------|---------------|--------|-------------------|---------------|-----------------|------------------------|
| Matic LP, et al. ([53]) | Carotid atheroma | Patients from the Karolinska Biobank | Transcriptomics (microarray platform, Illumina, Affymetrix) Proteomics (LC-MS/MS platform, Thermo) | CP, P | Systems biology | Identification of enriched pathways for carotid atheroma including cell proliferation, nitric oxide signaling, lipoprotein, and apoptotic particle clearance, immune cell activation, chemokine secretion, blood coagulation, and extracellular matrix disassembly were dominant in plaques by transcriptomics. Extracellular matrix, heme-binding, and platelet-derived growth factor binding were the most enriched functional categories by plaque proteomics. Integrative analysis showed BLVRB as the only significant candidate enriched both in plaques and plasma. | BLVRB, HMOX1 | In THP-1 macrophages iron stimulated an induction of BLVRB and HMOX1 was observed. |
| Lalowski MM, et al. ([54]) | Heart regeneration | Mice | Transcriptomics (mRNA sequencing, Illumina) Proteomics (LC/MS platform, Waters) Metabolomics (UPLC-MS/MS platform, Metabolon) | H | Systems biology | The decrease of the heart regeneration capacity was associated with a transition from fructose-induced glycolysis under hypoxic conditions to oxidative phosphorylation, with an increase in oxidative stress, suggesting a switch from hyperplasia to hypertrophy growth. Furthermore, they found enrichment of the glycolytic pathway, mTOR, plasmatas metabolism, methionine and histidine metabolism, lipid peroxidation, and sphingolipid signaling as novel pathways involved in heart regeneration. | Cpt I and II, Acsa2, Acsb1, Ecl1, Hadha, Hadhb, and Hsd17b10 | NA |
| Suhr K, et al. CAD ([35]) | KORA and TwinsUK cohorts, CARDioGRAM. | Genomics (microarray platform, Illumina, Affymetrix) Metabolomics (HPLC/MS platform, Metabolon) | B, P. | Genotype-dependent metabolic phenotypes | Some genetic loci that regulate blood metabolite concentrations were also associated with CAD risk (NAT2, ABO, CPS1, NAT8, ALPL, KLKB1). The biochemical function of the associated metabolic traits identified may support a possible role in heart disease. | NAT2 (1-methylxanthine/4-acetamidobutanoate); ABO (ADpS GEGDFXAEFGGVR/ADSGEGDFXAEFGGVR); CPS1 (Glycine); NAT8 (N-acetyltornithine); ALPL (ADpSSEGDFXAEFGGVR); KLKB1 bradykinin des-arg(9). | NA |
TABLE 1 | Continued

| References | Phenotypes | Population of study | Omic strategy | Tissue | Analysis strategy | Main findings | Genes involved | Functional confirmation |
|------------|------------|---------------------|---------------|--------|-------------------|---------------|----------------|------------------------|
| Feng Q, et al. (40) | CAD | 59 CAD patients and 43 healthy controls | Metabolomics (HPLC/MS platform, Thermo) | P | Association of metabolites with microbiome data | Some metabolites were significantly associated with gut microbiota and CAD risk (GlcNAc-6-P, mannitol, and 15 plasma cholines). Moreover, these identified metabolites show correlations with species of intestinal microbiota (Clostridium sp. and Streptococcus sp.). | LPCs, glycerophosphocholines, L-Arginine, GlcNAc-6-P, and paraxanthine | NA |
| Cui X, et al. (50) | Chronic heart failure | 53 CHF patients and 41 controls | Metabolomics (LC/MS platform, Thermo) | P | Correlation between changes in metabolites and gut microbiome associated with CHF | Enriched bacteria in CHF such as Veillonella were inversely correlated with cardiovascular protective metabolites such as niacin, cinnamic acid, and orotic acid. Furthermore, they found a positive correlation between the high sphingosine 1-phosphate levels and several CHF-enriched bacteria such as Veillonella, Coprobacillus, and Streptococcus. | Veillonella- niacin, cinnamic acid, and orotic acid | NA |
| Talukdar H, et al. (39) | CAD | GWAS of CARDioGRAMplusC4D and DIAGRAM studies. Mice (HMDP) | Genomics (microarray platform, Illumina, Affymetrix) | AAW, SF, VAF, LIV | Marker Set Enrichment Analysis (co-expression modules). Cross-species validation using the HMDP | Identification of 30 CAD-causal regulatory gene networks interconnected in vascular and metabolic tissues | POLR21, P2BP1, AIP, DRA1, MRPL28, PCBD1, ZNF91 | Validation of key divers in a THP-1 foam cells |
| Shu L, et al. (41) | T2D | GWAS data of five multi-ethnic studies including AA, EA, and HA. GWAS of CARDioGRAMplusC4D and DIAGRAM studies. Mice (HMDP) | Genomics (microarray platform, Illumina, Affymetrix) | 16 tissues including B, SF, ADR, ART, DT, IS, HY, LIV, LY, SHLM, TG, VE | Marker Set Enrichment Analysis (co-expression modules). Cross-species validation using cardiac-metabolic traits in the HMDP | Co-expression modules between CAD and T2D showed enrichment of pathways that regulate the metabolism of lipids, glucose, branched-chain amino acids, oxidation, extracellular matrix, immune response, and neuronal system. Identification of 15 key drivers associated with both CAD and T2D | ACAT2, ALCY, CAV1, COL6A2, COX7A2, DBI, HMGCR, ID1, IGF1, MCAM, MEST, MSTM1, PCOLCE, SPARC, and ZFP36 | SiRNA knockdown and in vivo knockout of CAV1 resulted in metabolic perturbations |

CAD, Cardiovascular Artery Disease; P, plasma; H, heart; B, blood; LIV, liver; AW, atherosclerotic arterial wall; MAM, atherosclerotic lesion-free internal mammary artery; ADR, atherosclerotic aortic root; SF, subcutaneous fat; VAF, visceral abdominal fat; SKLM, skeletal muscle; ADR, Adrenal gland; HCA, Human coronary artery smooth muscle cells; ART, Artery; DT, Digestive tract; IS, Islet; HY, Hypothalamus; LIV, Lymphocyte; TG, Thyroid gland; VE, Vascular endothelium; AV, Aortic valve; M, monocytes; CEC, Coronary endothelial cells; CR, Carotid plaque.

*For complete list of genes see reference.
and analysis of quantitative proteomics (87). Recently, Emilsson et al. showed that co-expression protein modules associated with complex diseases are highly regulated by cis and trans acting genetic variants (88). Therefore, the integration of proteomic data can add valuable information about the molecular processes involved in the development of CAD. One of the more interesting studies incorporating proteomic data in mice was conducted by Lau et al. which in addition to genomic and proteomic data, integrated protein dynamics (51). This study showed modules involved in cell adhesion, glycolytic process, actin filament organization, translation, and sodium ion transport associated with heart hypertrophy (51). In another multi-omics study conducted by Schlotter et al. for the identification of mechanisms involved in calcified aortic valve disease (CAVD) (52), the authors performed global transcriptomics and proteomics of human stenotic valves to identified novel regulatory networks in CAVD. Novel potential molecular drivers of CAVD development and progression were identified including alkaline phosphatase, apolipoprotein B, matrix metalloproteinase activation, and mitogen-activated protein kinase. Moreover, this approach also identified inflammation pathways as a significant contributor to CVD (52). This study emphasizes the relevance of extensive phenotypic characterization for multi-omics approaches to define markers associated with disease subgroups and to design more specific therapeutic strategies. In summary, these studies showed that the knowledge generated from the integration of genomics, epigenomics, transcriptomics, and proteomics could provide initial insights into the identification of mechanisms for cardiovascular diseases.

**METABOLOMICS AND METAGENOMIC STUDIES FOR THE STUDY OF CAD**

Metabolomics and metagenomics represent additional layers of complexity because they integrate the influences of the intake, utilization and flux of nutrients. Moreover, these omics data have proven to be useful tools for the identification of biomarkers with potential clinical applicability (89). However, studies integrating metabolomics, lipidomics, or metagenomics data in the context of CAD are limited (Table 1). In a GWAS study for metabolite levels conducted by Suhre et al. (35), the authors found several loci including ABO, NAT2, CPS1, NAT8, ALPL, KLKB1 genes associated with both metabolites and a high risk of CAD (35). Interestingly, KLKB1 was associated with bradykinin concentrations and with a higher CAD risk. It is known that bradykinin is a potent endothelium-dependent vasodilator that contributes to vasodilation and hypotension (90). These findings suggest that the integration of metabolomic data with other omic data can help to identify novel biomarkers for CVD diagnosis. Regarding studies integrating metagenomic data, there are only two studies for CVD so far that integrate metabolomics and metagenomics data (40, 50) (Table 1). These studies have shown species of bacteria associated with risk of CAD and plasma metabolites. For example, the bacteria Veillonella was associated with chronic heart failure and was also inversely correlated with known cardiovascular protective metabolites such as niacin, cinnamic acid and orotic acid (50). Nevertheless, it should be noted that these studies are only based on correlations and do not make an integrative analysis of the data, which reflects the complexity and the opportunity to develop novel statistical approaches.

**INTEGRATION OF MULTI-OMICS, MULTI-ETHNIC, AND MULTI-SPECIES MODELS OF DISEASE**

It has been suggested that comparison of “omics” data between human and animal models can provide an important contribution to the understanding of the molecular mechanism implicated in CAD (24). While studies in humans have greater translational potential, studies using animal models can help validate their biological relevance and to recapitulate the findings in humans under different environmental stimuli (22, 24, 78). This has been demonstrated in recent studies integrating multi-omics approaches for the study of CAD in both humans and animal models (39, 41). An example of a large-scale integrative multi-omic approach is the study conducted by Shu and colleagues that involved CAD and T2D GWAS data of five multi-ethnic studies (41). In this study, genetic and transcriptomic data of 16 relevant tissues for CAD were included to construct co-regulation networks for CVD and T2D (41). This network modeling allowed the identification of pathways involved in lipid metabolism, glucose, and branched-chain amino acids, along with process involved in oxidation, extracellular matrix, immune response, and neuronal system in CAD and T2D (41). Moreover, this strategy helped to dissect the molecular mechanism of HMGCR, identified as a top key driver for both CAD and T2D. Interestingly, the authors showed that HMGCR was associated with CVD and T2D in opposite directions, while genetic variants in HMGCR decrease CVD risk, they increase T2D risk. These findings could have important implications in the pharmacological treatment of both diseases. The integration of existing omics-data from mice and humans deposited in the cardiovascular disease database (CVDdb), including, microRNA, genomics, proteomics and metabolomics, has recently been analyzed to identified novel drivers for CVD. In an exercise to demonstrate the utility of the CVD database, integrative analysis of this “omics” studies showed enrichment of lipid metabolism, extracellular matrix remodeling, inflammation, and cardiac hypertrophy pathways. In addition, regulatory mechanisms mediated through miRNAs associated with the development of CAD were reported (47). Altogether, these studies illustrate that high-level integration approaches are powerful tools to extract robust biological signals across molecular layers, phenotypes, tissue types, and even species and to prioritize new therapeutic avenues for cardiometabolic diseases. Of note, there is a limited overlap in the metabolic regulators, co-expression modules and key driver gene identified across different multi-omics studies for CVD, except for markers involved in lipid metabolism which seem to be consistent among different studies. This highlights the importance of lipid
metabolism in the development of cardiovascular disorders (91–93). Discrepancies of these findings could be explained by differences in the statistical tools, phenotypic characterization, ethnic origin, sex, and pathophysiological conditions (13, 23–25, 79, 94).

DATA INTEGRATION USING FREELY AVAILABLE PUBLIC DATABASES

The access to big biologic public databases allows the integration of genomic data with other “omics” including transcriptomics, metabolomics, and proteomics datasets through freely available public databases such as GTEx (95) Encode (Encode project c, Roadmap (Roadmap Epigenomics Consortium, 2015), Snyderome (96) and bioRxiv, to mention a few. One of the main advantages of these databases is that allow simultaneous analysis of regulatory mechanism in different tissues, which are usually difficult to obtain in genetic studies conducted in humans. In this regard, the Genotype-Tissue Expression (GTEx) project is one of the most complete gene expression datasets currently available. This database was generated as a repository for identifying genetic variants associated with changes in gene expression (expression quantitative trait loci, eQTLs) and contains a broad tissue collection obtained from deceased donors. The last release v7, provides 11,688 transcriptomes from 714 individuals and 53 tissues. In addition GTEx also includes pathology and histology data as well as other characteristics as ethnicity, age, and sex (95). Moreover, in order to increase information about potential molecular mechanisms, the Enhancing GTEx (eGTEx) project extends the GTEx project to combine gene expression with DNase I hypersensitivity, ChiP–seq, DNA and RNA methylation, ASE, protein expression, somatic mutation, and telomere length assays (97). The Encyclopedia of DNA Elements (ENCODE) project has identified and annotated a significant amount of functional elements in the human and mice genome through diverse approaches as DNA hypersensitivity, DNA methylation, and immunoprecipitation (IP) assays of proteins that interact with DNA and RNA. The last version includes over 35 high-throughput experimental methods in > 250 different cell and tissue types, resulting in over 4,000 experiments. As GTEx database, ENCODE also includes relevant information about ethnicity, sex and age (98). Additional databases such as Roadmap (99), which has an extensive collection of DNA methylation, histone modifications, chromatin accessibility, and small RNA transcripts. The utility of these databases has been demonstrated in several studies for CAD, where their integration with genetic data facilitated the identification of regulatory mechanisms, potential targets and allows the functional validation. One example, is the prediction of the disruption of C/EBP binding site by the G allele of rs12740374 SNP using ENCODE data, functional studies showed that this variant results in a lower transcription of the SORT1 gene in liver and a higher VLDL-secretion, explaining the association of the variant with LDL-C levels in genetic studies (Figure 1) (74). Therefore, the integration of various data frameworks could be highly successfully to understand the mechanisms implicated in disease manifestation.

FUTURE DIRECTIONS

The identification of causal genes is a critical step toward the translation of genetic loci into biologic processes. The integration of “omic” strategies will accelerate the identification, in a more precise way, of novel molecular mechanisms implicated in CVD. This may eventually result in the characterization of novel pathways and drug targets. Although multi-omics approaches have been successfully applied for the investigation of cardiovascular diseases, the number of studies using this approach is still limited. These studies have been primarily focused on the integration of genomics, transcriptomics, epigenomics, and proteomics. Given the potential of metabolomics, metatranscriptomics, and metagenomics as tools for the identification of biomarkers with potential clinical applicability, the integration of such data will increase the understanding of cardiovascular diseases and accelerate the identification of new diagnostics or therapeutic targets (100). Finally, research efforts should be directed to the application of multi-omics and the generation of big data in more diverse populations and into the investigation of sex-specific mechanisms.

AUTHOR CONTRIBUTIONS

PL-M, JW, and AH-V drafted and edited the manuscript.

FUNDING

AH-V is funded by the NIH U54 DK120342 grant and NIH/CTSI UL1 TR00188. JW is funded by the NIH K08 HL133491 and NIH R01 HL129639.

REFERENCES

1. Writing Group M, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, et al. Heart disease and stroke statistics-2016 update: a report from the american heart association. Circulation. (2016) 133:e38–360. doi: 10.1161/CIR.0000000000000350
2. Zdankovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease in a study of twins. N Engl J Med. (1994) 330:1041–6. doi: 10.1056/NEJM199401143301503
3. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med. (1994) 330:1041–6. doi: 10.1056/NEJM199401143301503
4. Won HH, Natarajan P, Dobbin A, Jordan DM, Roussos P, Lage K, et al. Disproportionate contributions of select genomic compartments and cell types to genetic risk for coronary artery disease. PLoS Genet. (2015) 11:e1005622. doi: 10.1371/journal.pgen.1005622
5. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. (2013) 45:25–33. doi: 10.1038/ng.2480
Coronary Artery Disease Genetics. C, A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. (2011) 43:339–44. doi: 10.1038/ng.782

Klarin D, Damrauer SM, Cho K, Sun YV, Tesloch TM, Honerl J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet. (2018) 50:1514–23. doi: 10.1038/s41588-018-0222-9

Myocardial Infarction Genetics C, Kathiresan S, Voight BF, Purcell S, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. (2011) 43:333–8. doi: 10.1038/ng.784

Teslovich TM, Lal A, Wang B, Batzoglou S, Sidow A. Multi-omic tumour data integration of multi-omics data from mouse diversity panel highlights tissue-specific pathways and networks underlying sexual dimorphism and sex in humans. Nat Commun. (2011) 2:196–208. doi: 10.1038/ncomms12002

Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagle B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature. (2011) 477:54–60. doi: 10.1038/nature10354

Foroughi Asl H, Talukdar HA, Kindt AS, Jain RK, Ermel R, Ruusalepp A, et al. Expression quantitative trait loci acting across multiple tissues are enriched in inherited risk for coronary artery disease. Circ Cardiovasc Genet. (2015) 8:305–15. doi: 10.1161/CIRCGENETICS.114.000640

Braenne I, Civelek M, Vilne B, Di Narzo A, Johnson AD, Zhao Y, et al. Prediction of causal candidate genes in coronary artery disease loci. Arterioscler Thromb Vasc Biol. (2015) 35:2207–17. doi: 10.1161/ATVBAHA.115.306108

Franzen O, Ermel R, Cohain A, Akers NK, Ruusalepp A, et al. Cross-tissue regulatory gene networks in coronary artery disease. Cell Syst. (2016) 2:196–208. doi: 10.1016/j.cels.2016.02.002

Feng Q, Liu Z, Zhong S, Li R, Xia H, Jie Z, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with complex metabolic states. PLoS Genet. (2016) 12:e1005979.
63. Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, et al.
62. Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M,
58. Lau E, Wu JC. Omics, big data, and precision medicine
57. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al.
54. Lalowski MM, Bjork S, Finckenberg P, Soliymani R, Tarkia M, Calza G, et al.
55. Samani NJ, Braund PS, Erdmann J, Gotz A, Tomaszewski M, Linsel-
51. Lau E, Cao Q, Lam MPY, Wang J, Ng DCM, Bleakley BJ, et al. Integrated
45. Chen G, Yao C, Hwang SJ, Liu C, Song C, Huan T, et al. Integrated proteomic analysis of cardiovascular disease reveals novel protein quantitative trait loci. Circulation. (2016) 134:A18806.
44. Lempainen H, Braemne I, Michel T, Tragante V, Vilne B, Webb TR, et al. Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. Sci Rep. (2018) 8:8343. doi: 10.1038/s41598-018-20721-6
43. Fernandes M, Patel A, Hasi H. C/V Db: a multi-omics expression profiling database for a knowledge-driven approach in cardiovascular disease. (CVD). PLoS ONE. (2018) 13:e0207371. doi: 10.1371/journal.pone.0207371
42. Wang F, Xu CQ, He Q, Cui IP, Li XC. Wang D, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. Nat Genet. (2011) 43:345–9. doi: 10.1038/ng.783
41. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. (2007) 316:1488–91. doi: 10.1126/science.1142447
40. Wang F, Xu CQ, He Q, Cui IP, Li XC. Wang D, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. Nat Genet. (2011) 43:345–9. doi: 10.1038/ng.783
39. Buscher K, Ehninger G, Gupta P, Pramod AB, Wolf D, Tweet G, et al. Natural variation of macrophage activation as disease-relevant phenotype
predictive of inflammation and cancer survival. *Nat Commun.* (2017) 8:16041. doi: 10.1038/ncomms16041

82. Rau CD, Civelek M, Pan C, Lujs AJ. A suite of tools for biologists that improve accessibility and visualization of large systems genetics datasets: applications to the hybrid mouse diversity panel. *Methods Mol Biol.* (2017) 1488:153–88. doi: 10.1007/978-1-4939-6427-7_7

83. Rau CD, Romay MC, Tuterayan M, Wang H, Santolini M, Ren S, et al. Systems genetics approach identifies gene pathways and Adamts2 as drivers of isoproterenol-induced cardiac hypertrophy and cardiomyopathy in mice. *Cell Syst.* (2017) 4:121–8 e4. doi: 10.1016/j.cels.2016.10.016

84. Breitling LP, Yang RX, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet.* (2011) 88:450–7. doi: 10.1016/j.ajhg.2011.03.003

85. Fernandez-Sanles A, Sayols-Baixeras S, Curcio S, Subirana I, Marrugat J, Elosua R. DNA methylation and age-independent cardiovascular risk, an epigenome-wide approach: the REGICOR study (REgistre GIroni del COR). *Arterioscler Thromb Vasc Biol.* (2018) 38:645–52. doi: 10.1161/ATVBAHA.117.310340

86. Wang X, He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun.* (2018) 9:5380. doi: 10.1038/s41467-018-07746-1

87. Schubert OT, Rost HL, Collins BC, Rosenberger G, Aebersold R. Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc.* (2017) 12:1289–94. doi: 10.1038/nprot.2017.040

88. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson J, et al. Enabling precision cardiology through multiscale biology and systems medicine. *JACC Basic Transl Sci.* (2017) 2:311–27. doi: 10.1016/j.jacbts.2016.11.010

89. consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet.* (2013) 45:580–5. doi: 10.1038/ng.2653

90. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell.* (2012) 148:1293–307. doi: 10.1016/j.cell.2012.02.009

91. eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat Genet.* (2017) 49:1664–70. doi: 10.1038/ng.3969

92. Johnson CH, Ivanisevic J, Suizadak G. Metabolomics: beyond biomarkers and towards mechanisms. *Rev Rev Mol Cell Biol.* (2016) 17:451–9. doi: 10.1038/nrm.2016.25

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.