Cardiovascular disease risk and pathophysiology in South Asians: can longitudinal multi-omics shed light?

Yan V. Sun\textsuperscript{1,2}, Chang Liu\textsuperscript{1}, Lisa Staimez\textsuperscript{3}, Mohammed K. Ali\textsuperscript{3,4}, Howard Chang\textsuperscript{5}, Dimple Kondal\textsuperscript{6}, Shivani Patel\textsuperscript{3}, Dean Jones\textsuperscript{7}, Viswanathan Mohan\textsuperscript{8}, Nikhil Tandon\textsuperscript{9}, Dorairaj Prabhakaran\textsuperscript{6}, Arshed A. Quyyumi\textsuperscript{10}, K. M. Venkat Narayan\textsuperscript{3}, Anurag Agrawal\textsuperscript{11}

\textsuperscript{1}Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, 30322, USA
\textsuperscript{2}Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, 30322, USA
\textsuperscript{3}Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, 30322, USA
\textsuperscript{4}Department of Family and Preventive Medicine, School of Medicine, Emory University, Atlanta, GA, 30322, USA
\textsuperscript{5}Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, 30322, USA
\textsuperscript{6}Public Health Foundation of India, New Delhi, India
\textsuperscript{7}Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine, School of Medicine, Emory University, Atlanta, GA, 30322, USA
\textsuperscript{8}Madras Diabetes Research Foundation, Chennai, India
\textsuperscript{9}All India Institute of Medical Sciences, New Delhi, India
\textsuperscript{10}Department of Medicine, Division of Cardiology, Emory University School of Medicine, Atlanta, GA, 30322, USA
\textsuperscript{11}Institute of Genomics and Integrative Biology, Council of Scientific and Industrial Research, New Delhi, India

Abstract

Cardiovascular disease (CVD) is the leading cause of mortality in South Asia, with rapidly increasing prevalence of hypertension, type 2 diabetes (T2DM) and hyperlipidemia over the last two decades. Atherosclerotic CVD (ASCVD) affects South Asians earlier in life and at lower body weights, which is not fully explained by differential burden of conventional risk factors. Heart failure (HF) is a complex clinical syndrome of heterogeneous structural phenotypes including two major clinical subtypes, HF with preserved (HFrEF) and reduced ejection fraction (HFrEF). The prevalence of HF in South Asians is also rising with other metabolic diseases, and HFrEF develops at younger age and leaner body mass index in South Asians than in Whites. Recent genome-wide association studies, epigenome-wide association studies and metabolomic studies of ASCVD and HF have identified genes, metabolites and pathways associated with CVD traits. However, these findings were mostly driven by samples of European ancestry, which may not accurately represent the CVD risk at the molecular level, and the unique risk profile of CVD in South Asians.
Such bias, while formulating hypothesis-driven research studies, risks missing important causal or predictive factors unique to South Asians. Importantly, a longitudinal design of multi-omic markers can capture the life-course risk and natural history related to CVD, and partially disentangle putative causal relationship between risk factors, multi-omic markers and subclinical and clinical ASCVD and HF. In conclusion, combining high-resolution untargeted metabolomics with epigenomics of rigorous, longitudinal design will provide comprehensive unbiased molecular characterization of subclinical and clinical CVD among South Asians. A thorough understanding of CVD-associated metabolomic profiles, together with advances in epigenomics and genomics, will lead to more accurate estimates of CVD progression and stimulate new strategies for improving cardiovascular health.

**Keywords**

multi-omics, heart failure, atherosclerosis, subclinical CVD, HFpEF, HFrEF, South Asians, diabetes

This article is included in the Wellcome Trust/DBT India Alliance gateway.
In the revised version, we added details of the current knowledge on heart failure subtypes, more recent omics studies of cardiovascular disease risk factors, as well as a review of proteomics and gene expression studies. We also further discussed the strategies and challenges of omics studies. A new summary table (Table 1) for recent omics studies of cardiovascular disease is included.

**Any further responses from the reviewers can be found at the end of the article.**

**Introduction**

Cardiovascular disease (CVD) remains the most important cause of mortality worldwide. CVD is now the leading cause of mortality in India, accounting for 25% of deaths\(^1\)\(^2\). The prevalence of CVD risk factors including hypertension, type 2 diabetes (T2DM), and lipid levels have increased rapidly over the last two decades\(^3\)\(^-\)\(^7\). Atherosclerotic CVD (ASCVD) phenotypes are heterogeneous\(^8\)\(^-\)\(^10\). The presence of subclinical ASCVD precursors like vascular dysfunction can be measured as endothelial dysfunction, arterial stiffness, or microvascular dysfunction. Subclinical structural changes can be detected by carotid intima-media thickening (CIMT), plaque deposition, coronary artery calcium (CAC), and reduced ankle-brachial index (ABI), all of which can predict future ASCVD events\(^11\)\(^-\)\(^14\). ASCVD affects South Asians earlier in life, with 52% of CVD deaths in individuals <70 years compared to 23% in the West, disparities that are not fully explained by differences in conventional risk factor burden\(^15\)\(^-\)\(^17\).

Heart failure (HF) is a complex clinical syndrome that results from structural and functional impairment of ventricular filling or output, and manifests itself in heterogeneous structural phenotypes (HF with preserved [HFrEF] or reduced ejection fraction [HFrEF]). HF prevalence in the US is projected to increase 46% from 2012 to 2030, resulting in over 8 million adults (≥18 years) with HF\(^18\). The prevalence of HF in South Asians is also rising with other metabolic diseases\(^19\)\(^-\)\(^21\), and HFrEF develops at younger age and leaner BMI than for Whites\(^22\).

Nearly half of the HF patients have HFrEF, with >90% being >60 years old, with rapidly increasing numbers\(^23\)\(^-\)\(^27\). Although numerous risk factors for HFrEF have been identified including hypertension, older age, female sex, obesity, diabetes, and renal disease\(^28\)\(^-\)\(^29\), there are currently no class I guideline recommended treatments for HFrEF that improve mortality\(^30\)\(^-\)\(^31\). HFrEF has received increased attention since HFrEF patients frequently experience delayed diagnosis and have limited treatment options. Recent studies have shown that HFrEF is a clinically heterogeneous disorder, consisting of subgroups with related comorbidities and pathophysologies, which lead to different progression trajectories\(^32\)\(^-\)\(^33\). Employing latent class analytics and clustering techniques based on widely available clinical variables, several investigators have shown that patients with HFrEF can be divided into distinct classes with differing outcomes\(^34\)\(^-\)\(^35\).

CVD progression can be affected by both gene and environment via different molecular pathways and mechanisms. Although the high throughput technologies have enabled accurate and cost-effective genotyping in large population samples, comparable high throughput measurements of environmental exposures are needed for large population studies. Gene-environment interaction is a common mechanism to explain complex disease risk, and inter-individual variability. Better understanding of gene-environment interactions and their causal relationships will point to pathways and mechanisms as potential targets for treatment and intervention. The gene-environment interaction study may also help understand the CVD risk among immigrant populations exposure to different environmental and lifestyle factors.

In this paper, we focus on the current progress of omics studies on ASCVD and HF, as well as the anticipation of implications among South Asian population. A summary of selected CVD omics studies listed are shown in Table 1.

**Genetic basis of atherosclerotic cardiovascular disease and heart failure**

Genome-wide association studies (GWAS) have identified a large number of genetic loci associated with coronary heart disease\(^36\)\(^-\)\(^38\), ASCVD\(^39\)\(^-\)\(^41\), HF\(^42\)\(^-\)\(^45\), and their risk factors, such as BMI\(^46\)\(^-\)\(^47\), blood lipids\(^48\)\(^-\)\(^50\), blood pressure\(^51\)\(^-\)\(^52\), and T2DM\(^53\); however, these genetic variations explain only a small portion of risk in populations\(^54\)\(^-\)\(^56\). Joint genetic-environmental effects may be a key mechanism responsible for unexplained CVD risk. For example, environmental exposures can modify the gene expression levels through epigenetic mechanisms and this epigenetic modification can be inherited across cell generations to exert a long-term impact on the development of CVD.

A large number of genetic associations have been identified in large population studies for CVD and risk factors. Although identified genetic associations have small effect sizes individually, polygenic risk score (PRS) can combine such individual effects into a much stronger predictor of a disease trait. A few studies have shown the successful identifications of CAD and relevant traits\(^57\)\(^-\)\(^59\). Earlier studies of PRS showed that the European ancestry-based can be transferred in other ancestry groups including South Asians, the associations of European ancestry-derived PRS were typically weaker in non-European ancestries. With large GWAS results from multiple ancestry groups, the PRS can be optimized to present ancestry-specific genetic risk among immigrant populations exposure to different environmental and lifestyle factors. Better understanding of gene-environment interactions and their causal relationships will point to pathways and mechanisms as potential targets for treatment and intervention. The gene-environment interaction study may also help understand the CVD risk among immigrant populations exposure to different environmental and lifestyle factors.
<table>
<thead>
<tr>
<th>PMID</th>
<th>First Author</th>
<th>Year</th>
<th>Phenotype</th>
<th>Sample size</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic basis of atherosclerotic cardiovascular disease and heart failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33532862</td>
<td>Hartiala JA</td>
<td>2021</td>
<td>Myocardial infarction</td>
<td>Discovery: 61,000 cases, 577,000 controls</td>
<td>European, East Asian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Replication: 165,000+27,000 subjects</td>
<td></td>
</tr>
<tr>
<td>33020668</td>
<td>Koyama S</td>
<td>2020</td>
<td>CAD</td>
<td>121,234 cases, 527,824 controls</td>
<td>East Asian, European</td>
</tr>
<tr>
<td>N/A</td>
<td>Assimes T</td>
<td>2021</td>
<td>CAD</td>
<td>~250,000 cases</td>
<td>European, African American, Hispanic, East Asian</td>
</tr>
<tr>
<td>26343387</td>
<td>Nikpay M</td>
<td>2015</td>
<td>CAD</td>
<td>60,801 cases, 123,504 controls</td>
<td>European, South Asian, East Asian, Hispanic, African American</td>
</tr>
<tr>
<td>29212778</td>
<td>van der Harst P</td>
<td>2018</td>
<td>CAD</td>
<td>Discovery: 34,541 cases, 261,984 controls</td>
<td>European</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Replication: 88,192 cases, 162,544 controls</td>
<td></td>
</tr>
<tr>
<td>31160810</td>
<td>Malik R</td>
<td>2018</td>
<td>Stroke</td>
<td>67,162 cases, 454,450 controls</td>
<td>European, East Asian, African, South Asian, mixed Asian, Latin American</td>
</tr>
<tr>
<td>31919418</td>
<td>Shah S</td>
<td>2020</td>
<td>HF</td>
<td>47,309 cases, 930,014 controls</td>
<td>European</td>
</tr>
<tr>
<td>23202125</td>
<td>Consortium CAD</td>
<td>2013</td>
<td>CAD</td>
<td>63,746 cases, 130,681 controls</td>
<td>European, South Asian</td>
</tr>
<tr>
<td>Epigenetic signatures identify molecular pathways of CVD that are activated in the context of environmental risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31424985</td>
<td>Agha G</td>
<td>2019</td>
<td>Myocardial infarction, CAD</td>
<td>11,461 subjects</td>
<td>European, African American</td>
</tr>
<tr>
<td>28515798</td>
<td>Nakatochi M</td>
<td>2017</td>
<td>Myocardial infarction</td>
<td>192 cases, 192 controls</td>
<td>Japanese</td>
</tr>
<tr>
<td>28172975</td>
<td>Rask-Andersen M</td>
<td>2016</td>
<td>Myocardial infarction</td>
<td>729 subjects</td>
<td>European</td>
</tr>
<tr>
<td>31615550</td>
<td>Westerman K</td>
<td>2019</td>
<td>CVD</td>
<td>Discovery: 2,023 subjects Validation: 2,587 subjects</td>
<td>European, African American, Hispanic</td>
</tr>
<tr>
<td>28838933</td>
<td>Meder B</td>
<td>2017</td>
<td>HF</td>
<td>Discovery: 41 cases, 31+31 controls Replication 1: 18+9 cases, 8+28 controls</td>
<td>European</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Replication 2: 82 cases, 109 controls</td>
<td></td>
</tr>
<tr>
<td>Proteomics and gene expression reveal the interactions between cell processes and external environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30587458</td>
<td>Bom MJ</td>
<td>2019</td>
<td>Coronary plaque morphology</td>
<td>196 subjects</td>
<td>European</td>
</tr>
<tr>
<td>32808014</td>
<td>Hoogeveen RM</td>
<td>2020</td>
<td>CVD</td>
<td>Discovery: 822 subjects Validation: 702 subjects</td>
<td>European</td>
</tr>
<tr>
<td>PMID</td>
<td>First Author</td>
<td>Year</td>
<td>Phenotype</td>
<td>Sample size</td>
<td>Ethnicity</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------</td>
<td>------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>25881932</td>
<td>Cheng ML</td>
<td>2015</td>
<td>HF</td>
<td>183 cases, 51 controls</td>
<td>East Asian</td>
</tr>
<tr>
<td>31092001</td>
<td>Wang Z</td>
<td>2019</td>
<td>CAD</td>
<td>3,598 subjects</td>
<td>African American, European</td>
</tr>
<tr>
<td>29838382</td>
<td>Bhupathiraju SN</td>
<td>2018</td>
<td>CAD</td>
<td>145</td>
<td>South Asian</td>
</tr>
<tr>
<td>23788672</td>
<td>Zheng Y</td>
<td>2013</td>
<td>HF</td>
<td>1,744 subjects</td>
<td>African American, European</td>
</tr>
<tr>
<td>29967392</td>
<td>Lanfear DE</td>
<td>2017</td>
<td>HF</td>
<td>516 subjects</td>
<td>European, African American</td>
</tr>
<tr>
<td>29967392</td>
<td>Andersson C</td>
<td>2019</td>
<td>HF</td>
<td>8,772 subjects</td>
<td>European</td>
</tr>
<tr>
<td>3366805</td>
<td>Palou-Marquez G</td>
<td>2021</td>
<td>CVD</td>
<td>2,015 subjects</td>
<td>European</td>
</tr>
<tr>
<td>3366805</td>
<td></td>
<td></td>
<td></td>
<td>914 subjects</td>
<td></td>
</tr>
</tbody>
</table>

**The multi-omics approach is critical to understand CVD risk at the molecular level.**
showing the benefits of including diverse ancestries in discovering novel genetic loci associated with diseases\textsuperscript{66}, and improving the variance explained across ethnic groups. The South Asian population, who constitute over 20\% of humanity, remain under-represented in large genomic, epigenomic and other -omic studies. More GWAS in South Asians would improve both ethnicity-specific, and trans-ethnic discoveries of ASCVD- and HF-related loci, thus, address the high burden of these diseases in this large and growing population.

**Epigenetic signatures identify molecular pathways of CVD that are activated in the context of environmental risk**

Epigenetics refers to molecular modifications that are unrelated to the primary DNA sequence and that can arise from environmental exposures\textsuperscript{66}. Epigenetic modification, through DNA methylation (DNAm) and other molecular mechanisms can regulate gene expression levels that can influence susceptibility to disease development\textsuperscript{65}, including atherosclerosis\textsuperscript{68} and chronic inflammation\textsuperscript{69}, two important pathophysiological processes leading to CVD. Furthermore, epigenetic markers are modified by age\textsuperscript{70,71} and environmental risk factors, such as smoking\textsuperscript{72} and poor nutrition\textsuperscript{73}. Thus, the epigenetic profile can capture many of the cumulative environmental effects that influence susceptibility to CVD and its adverse outcomes. An epigenome-wide association study (EWAS) is a study of epigenetic markers across the entire genome in a population exhibiting a specific trait\textsuperscript{74}. EWAS provides an unbiased approach for identifying molecular mediators of genetic and environmental factors that may explain residual risk of disease\textsuperscript{75} and has been applied to investigate CVD\textsuperscript{76–79} and risk factors\textsuperscript{80–83}.

A recent EWAS of 11,461 individuals of European or African ancestry identified 52 DNAm sites significantly associated with incident coronary artery disease (CAD; n=1,895) using peripheral blood samples\textsuperscript{82}. This robust study reported epigenetic associations with effect sizes of a clinically relevant magnitude. Another EWAS identified 59 DNAm sites that associated with risk of dilated cardiomyopathy in left ventricular and peripheral blood samples\textsuperscript{79}. Bioinformatic analyses of differentially methylated regions showed enrichment for modification around binding sites for transcription factors involved in cardiac phenotypes. Studies in large populations have also identified DNAm sites associated with traditional CVD risk factors including age\textsuperscript{71,76–78}, BMI\textsuperscript{80,89}, diabetes\textsuperscript{90,91,92}, blood lipids\textsuperscript{93}, smoking\textsuperscript{72,94} and inflammation\textsuperscript{95}. These DNAm markers may provide a measure of longitudinal CVD risk\textsuperscript{96}, independently predicting future CVD events. However, DNAm sites need to be rigorously examined across genetically distinct cohorts. The study of incident T2DM conducted among Indian Asians and Europeans showed a 2.5 times higher adjusted risk among Indian Asians than Europeans, and five loci including ABCG1, PHOSPHO1, SOCS3, SREBF1, and TXNIP were associated with incident T2DM among Indian Asians and replicated among Europeans\textsuperscript{81}. The other study of incident T2DM reported additional loci such as PHGDH and CPT1A, which were discovered among Europeans and replicated among Indian Asians\textsuperscript{81}. The study of blood pressure compared the methylation profiles between Europeans and South Asians revealed many distinct loci between the two ancestries with a small overlap\textsuperscript{84}. To our current knowledge, some known trans-ethnic epigenetic loci might be transportable from Europeans to South Asians, but there is still a lack in studies specifically among large South Asian populations to further explore novel loci to explain the higher risk. A number of EWAS have examined DNAm patterns associated with air pollution, which is a known environmental risk factor for CVD\textsuperscript{87,88}. Since epigenomic profile can be modified due to changes in environmental and socio-behavioral factors, a longitudinal study can reveal the dynamics of the epigenomic profile, and potential causal or mediation effects in relation of CVD risk and progression. In sum, epigenetic profiles provide a signature of the cumulative burden of life-long exposure to CVD risks.

**Proteomics and gene expression reveal the interactions between cell processes and external environment**

The proteomics technologies have evolved rapidly in the past two decades. Recent applications of proteomics in population studies have produced interesting findings in CVD research\textsuperscript{89}. A recent study of the healthy human heart tissue collected from autopsy determined the healthy heart proteome. The resulted database, which included over 10,700 proteins, is a comprehensive resource for the downstream investigations\textsuperscript{90}. In a targeted proteomics study, two protein signatures for high-risk plaques and absence of coronary atherosclerosis were identified among a cohort with suspected coronary artery disease\textsuperscript{91}. The prediction accuracy of a model constructed by 50 proteins showed a better prediction accuracy in adverse cardiovascular events than a traditional risk factor model\textsuperscript{102}. These findings are critical in CVD risk prediction and differentiation of CVD subtypes. In addition, gene expression profiling in CVD also enables a better understanding of pathophysiology of CVD\textsuperscript{103}. Patterns of gene expression of human aorta tissue was investigated to identify genes with prediction power in atherosclerosis\textsuperscript{104}. Studies of non-coding transcriptome, which even though have limited protein-coding functions, are discovering the pathology of CVD. To eventually realize precision medicine, we are still facing the challenges of standardization of methodologies and translation\textsuperscript{105}. In addition, researches in this area largely depend on the availability of tissues from autopsy, thus the clinical translation and implementation has been limited. However, such studies facilitate a better understanding on the disease causal pathway and mechanism. Particularly, there has been a lack of similar researches particularly among South Asians, which is urging the future efforts of investigation.

**Metabolomic signatures present complex metabolic state and environmental exposure**

The metabolome is a global identification of all small molecules produced by cells during metabolism or obtained from environmental exposures. The metabolome thus provides a direct functional readout of cellular activity and physiologic status that can potentially be used for early disease identification, study of treatment effects, and for prognostication of disease progression\textsuperscript{106–108}. It reflects the combined systemic effects of genetic, lifestyle, and environmental factors. Metabolomics is
an emerging discipline that has the potential to transform the study of biological responses to environment exposures that underlie disease development. The untargeted metabolomic approach provides unbiased coverage of metabolites with greater breadth than targeted methods. Advances in untargeted metabolomics have increased the number of metabolites analyzed, thereby improving accuracy of disease detection and quantification. Metabolomics holds promise in elucidating interactions between genes and environment (such as air pollution) to uncover the pathophysiology and underlying molecular pathways of complex disorders such as ASCVD and HF.

Metabolomic research in humans has shown that modification or dysregulation of numerous metabolites, including amino acids, phospholipids, short-chain acylcarnitines, and nitric oxide synthesis, are associated with CVD risk and outcomes. In a study of 2,232 African and 1,366 European Americans from the Atherosclerosis Risk in Communities (ARIC) study (633 incident CAD cases), 19 metabolites collectively improved CAD risk prediction. Metabolomics study of dietary patterns among Asian Indians were found associated with cardiometabolic risk. A study of 145 Asian Indians in the Metabolic Syndrome and Atherosclerosis in South Asians Living in America (MASALA) pilot study revealed the metabolite pattern of branched-chain amino acids, aromatic amino acids, and short-chain acylcarnitines, which are representative of a “Western/ nonvegetarian” dietary pattern associated with adverse cardiometabolic profile. Recent studies also reported changes in global metabolism in relation to HF risk and outcomes. A study of 515 HF patients identified a panel of metabolites that improved prediction of HF-related mortality and re-hospitalization, with variation between HF subgroups. By demonstrating diagnostic and prognostic value in HF risk, these studies suggest that metabolomic research will distinguish HF subtypes.

**The multi-omics approach is critical to understand CVD risk at the molecular level**

Genetics is one of the primary sources of epigenetic and metabolic variation. Recent GWAS have identified >160 genomic loci for CAD, 11 loci for HF, and >300 loci for T2DM. However, biological functions of most identified genetic loci remain unknown. Therefore, assessment of the functional linkage between identified epigenetic makers, metabolites, and genetic variants would be fruitful. Unlike the genome profile, epigenomic, transcriptomic, proteomic and metabolomic profiles can be modified by environmental exposures, physiological conditions and disease status. Genomic data will complement epigenomic and metabolomic markers by identifying complex biological processes at the systems levels. Current omics studies predominantly use a cross-sectional design, which enables the novel biomarker discovery but limited to infer causal associations due to confounding and reverse causation. A longitudinal design including baseline omics and incident CVD would better demonstrate the prediction utility of omics markers for CVD progression, and has been implemented in omic association studies in other ancestry groups. However, such a design doesn’t incorporate the longitudinal changes of omics markers in relation to varying environment, pathobiology and disease progression. Repeated measurements of omics profiles excluding genomics would be important to understand long-term risk and natural history of chronic diseases such as CVD. Such omics changes may also help identify reversible targets for novel interventions for CVD outcomes.

A recent study using longitudinal big data including genome, immunome, transcriptome, proteome, metabolome, microbiome and wearable monitoring have showed the potential of revealing cardiovascular pathophysiology on a molecular basis. Although increasing number of population studies have measured multi-omics data, recent studies have focused on single -omic association study and used additional -omics data to better understand the molecular functions of identified associations. Integrated multi-omics studies of CVD outcomes have been limited. A large HF study of over 8,000 participants within the Framingham Heart Study utilized integrative trans-omics data including genetic variations, DNA methylation, and gene expression data to reveal genetic contributions towards HF. The transportability of such findings to South Asians needs to be evaluated. Another study integrating DNA methylation and gene expression data identified independent latent factors associated with CVD. The unsupervised machine learning of multi-omics successfully improved classification and discrimination. Additionally, our recent joint epigenomics-metabolomics study of smoking demonstrated that these multi-omic layers capture complementary components of biological systems in response to widespread risk exposures such as air pollution. By jointly analyzing genomic, epigenomic, and metabolomic profiles, we hypothesize that it could lead to identification of key genes and pathways that may be the molecular mediators of CVD. The molecular functions of identified epigenetic and metabolic markers, key genes and pathways involved in subclinical and clinical CVD will help us develop future targeted studies to improve comorbid disease prevention and clinical care strategies.

To extend this understanding to high-risk populations such as South Asians, it is important that such studies be performed in longitudinal cohorts that adequately represent the ethnicity and the environment. A longitudinal design enables the detection of changes in the characteristics of the target population at both the group and the individual level. Since the multi-omics profile (e.g., epigenome and metabolome) can be modified by dynamic and cumulative environmental factors, capturing the longitudinal multi-omics would benefit the understanding of risk development and progression of subclinical and clinical CVD in South Asians. Longitudinal changes of individual omic profile related to disease and aging have been documented in humans, primates and mice. Longitudinal design can also distinguish cause from consequence using appropriate modeling. With large sample size, properly implemented mediation analysis and strong genetic instrumental variables for Mendelian Randomization analysis, future longitudinal studies in South Asian populations can further control for...
genetic and environmental confounders to better address causal relationship between -omic markers, environmental factors and CVD.

**Strategies of omics studies**

Due to the rapidly growing technologies such as next generation sequencing and untargeted metabolomics, massive amount of accurate data can be obtained in a cost-effective manner. Therefore, there has been an urgent need for new methodologies and algorithms to form, for the purposes of computationally intensive data managing and analysis[16]. For GWAS, tools such as PLINK[27], RVTESTS[28] and GENESIS[29] now allow the incorporation of many data pieces such as genotypes, annotations, allele frequencies and phenotypes to be processed efficiently in particular formats. For untargeted MWAS, mixOmics[30] and xmsPANDA (https://github.com/kuppal2/xmsPANDA) has been utilized to conduct feature selection, and the software Mummichop[31] for pathway exploration based on untargeted metabolomics data has been made available to bypass the feature identification stage. In addition, various tools have been developed for network analysis. For example, the weighted gene co-expression network analysis can be implemented using WGCNA[32] to identify clustering of genes, the software xMWAS[33] can perform the integration and differential network analysis for multiple layers of omics data. The evolving omics researches of cardiovascular risk are coupled with these tools to achieve better understanding of the disease biological mechanism and pathway.

**The multi-omics approach may address several challenges in CVD research facing South Asians**

South Asians are an understudied population at high CVD risk even at low body weight and young ages, with a high propensity to diabetes, dyslipidemia, and hepatic steatosis, features that may provide new insights into CVD risk in the presence and absence of traditional risk factors. Given the genetic and molecular underpinning of CVD and risk factors, the genetic and molecular markers identified from multi-omics research can help accurately profile the CVD risk, given the unique characteristics among South Asians. Secondly, the multi-omics approach also holds the promise of revealing the heterogeneous mechanisms of CVD and risk factors. For example, HFpEF patients can be clustered into distinct subclasses with different clinical outcome using available demographic and clinical variables[30,41]. T2DM is also a multi-factorial disease that involves numerous genetic pathways and many environmental factors[14], and has high prevalence in South Asian populations. Recent studies of T2DM subgroups have revealed striking heterogeneity of T2DM risk, etiology and outcomes, which may be further illustrated by multi-omic profiling. Lastly, multi-omics can help explain inter-individual variability in response to environmental risk exposures. Exposures to environmental risk such as air pollution (common and severer CVD risk in South Asia) are multifactorial and time-varying, thus, their measurements can be incomplete, costly, and imprecise for pathophysiological effects.

Omic technologies such as epigenomics and metabolomics represent exogenous and endogenous effects related to CVD risk and have emerged as key components of exposure measures[35,16]. They can capture the biological response to environmental exposures, thus, providing more precise risk assessment for subclinical and clinical CVD. Improved omics measurements of environmental exposure will also enable large scale gene-environment interaction study to further uncover molecular mechanisms underlying CVD pathophysiology.

In addition to common genomic and multi-omics factors of CVD across ancestry groups, studying multi-omics among large population samples within South Asians may also discover genetic variants or molecular signatures unique for South Asians. Incorporating these multi-omic profiles can optimize the diagnosis, treatment and prognosis of CVD, which is the most important and still growing health burden for South Asian populations.

**Conclusion**

Despite the available tools and fruitful research conducted among European ancestry, the implementation of similar research among South Asians remains challenging. Large high-quality studies are needed, with the requirements of high volume recruiting of population-based study samples, well-defined disease and phenotypes, long time follow-up, establishment of data registries, close collaborations between scientists with various expertise, such as study design, molecular biology and bioinformatics. There is a growing effort of research among the South Asian population, such as the Center for cArdio-metabolic Risk Reduction in South Asia (CARRS) study[37], which is a large longitudinal study of 28,000 subjects across three large cities in South Asia. Studies that employ the integrative multi-omic approach to disentangling complex molecular systems underlying CVD are still anticipated. A thorough understanding of CVD-associated metabolomic profiles, together with advances in epigenomics and genomics, will lead to more accurate estimates of CVD progression and stimulate new strategies for improving cardiovascular health. Combining high-resolution untargeted metabolomics, epigenomics with a rigorous, longitudinal design would provide comprehensive molecular characterization of subclinical and clinical CVD among South Asians – a large global population with unique CVD patterns, and with potential variations in phenotypes. To fully understand the gene-environment interplay, it would also be useful to have studies of South Asians living in South Asia as well as South Asian emigrants.

**Data availability**

No data are associated with this article.

**Acknowledgments**

The authors thank the staff and participants of the CARRS study for their important contributions.
References


Wellcome Open Research 2021, 5:255 Last updated: 20 NOV 2021
44. Assimes T, Tcheandjieu C, Zhu X: A large-scale multi-ethnic genome-wide association study for coronary artery disease. PREPRINT (Version 2) available at Research Square. Publisher Full Text


82. Dick KJ, Nelson CP, Tasprouni L, et al.: DNA methylation and body-mass index:
PubMed Abstract | Publisher Full Text | Free Full Text

133. Uppal K, Ma C, Go YM, et al.: xMWAS: a data-driven integration and
PubMed Abstract | Publisher Full Text | Free Full Text

heterogeneity within and between different diabetes types. Diabetologia.
PubMed Abstract | Publisher Full Text

135. Wild CP, Scalbert A, Herceg Z: Measuring the exposome: a powerful basis for
evaluating environmental exposures and cancer risk. Environ Mol Mutagen.
PubMed Abstract | Publisher Full Text

136. Miller GW, Jones DP: The nature of nurture: refining the definition of the
PubMed Abstract | Publisher Full Text | Free Full Text

to assess burdens from multiple perspectives. BMC Public Health. 2012; 12:
701.
PubMed Abstract | Publisher Full Text | Free Full Text
Chunyu Liu
Department of Biostatistics, Boston University, Boston, MA, USA

I believe the authors have adequately addressed my comments. Thank you.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical genetics and genomics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Nilay Shah
Division of Cardiology/Department of Medicine and Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

The authors have satisfactorily revised the article in response to reviewer comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cardiovascular disease epidemiology, South Asian cardiovascular health and disease prevention.
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 09 December 2020

https://doi.org/10.21956/wellcomeopenres.17960.r41637

© 2020 Shah N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nilay Shah
Division of Cardiology/Department of Medicine and Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

In the review article “Cardiovascular disease risk and pathophysiology in South Asians: can longitudinal multi-omics shed light?” by Sun et al., the authors provide a summary of research findings from genetic, epigenetic, and metabolomic studies of cardiovascular risk factors and cardiovascular disease. The narrative begins to detail how these “multi-omic” approaches may be used to better understand the mechanisms of cardiovascular diseases, and concludes with a recommendation to increase “multi-omics” research in the South Asian population, with particular emphasis on longitudinal epigenomics and metabolomics to comprehensively characterize subclinical and clinical cardiovascular disease at the molecular level.

The authors have identified and described an important and growing field of research that is relevant to South Asians, a population that is disproportionately affected by cardiovascular diseases. A number of comments and suggests are submitted for the authors’ consideration to strengthen the relevance, clarity, and impact of their review:

1. The narrative review of “multi-omics” provided by the authors is superficial, and does not sufficiently detail these growing methodologies. The authors are encouraged to broaden their narrative of these -omics tools, including recent findings and how these tools are being used to develop strategies to improve cardiovascular health in European ancestry populations (which the authors acknowledge are the populations for which most data are available). This broader narrative and contextualization would more comprehensively describe the current state of this field, and help the reader understand how -omics approaches may be applied in South Asians.

For instance, recent GWAS have been conducted in South Asian diaspora populations, and have contributed to a variety of strategies to understand cardiovascular risk in this population such as polygenic risk scoring (e.g., Wang M et al. “Validation of a genome-wide polygenic risk score for coronary artery disease in South Asians.” J Am Coll Cardiol. 2020;76(6):703-14.) This and other recent similar studies would be helpful to include.
2. Since the narrative review is cursory, it is not clear how the authors envision “multi-omic” methodologies should be deployed in South Asian populations. What are the key research questions with these methods in the South Asian population? What do the authors hypothesize will be found? Are there any characteristics specific to the South Asian population that should be accounted for (for example, emerging epidemiologic work identifies South Asian subgroups, e.g. Pakistani, Indian, etc.; how does this fit into research evaluating environmental influences on epigenomics or metabolomics)? Addressing these questions seems important for this review to understand how longitudinal multi-omics might “shed light” on CVD in South Asians.

3. Are longitudinal -omics a component of research in other (including European ancestry) populations? What insights have been gained from such work? The authors suggest a longitudinal -omics approach in their concluding paragraphs, but have inadequately supported the potential merits of this strategy.

4. This review would be a prime opportunity to move this field forward beyond only stating that these -omics tools should be applied in South Asian populations. What are the existing challenges/barriers preventing these tools from being implemented in the study of South Asian CVD and CVD risk? (In parallel, why have most studies using -omics approaches occurred in European ancestry populations?) What strategies would help overcome these barriers? The authors are recommended to consider addressing these questions in their review to provide important context to the state of the science and a roadmap to achieve their recommended future directions.

5. The authors’ concluding paragraph introduces new concepts that were not previously discussed (for example, “HFpEF heterogeneity,” and the differentiation of South Asians in South Asia versus diaspora populations, which certainly would seem relevant with respect to environment-gene interaction). Consequently, their conclusions seem to be obscured and do not follow from their narrative. I recommended discussing these components of research gaps in South Asians earlier in the review, in order to justify their conclusions.

References

Is the topic of the review discussed comprehensively in the context of the current literature?
No

Are all factual statements correct and adequately supported by citations?
Yes

Is the review written in accessible language?
Yes
Are the conclusions drawn appropriate in the context of the current research literature? Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Cardiovascular disease epidemiology, South Asian cardiovascular health and disease prevention

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 07 May 2021

Yan Sun, Rollins School of Public Health, Emory University, Atlanta, USA

1. The narrative review of “multi-omics” provided by the authors is superficial, and does not sufficiently detail these growing methodologies. The authors are encouraged to broaden their narrative of these -omics tools, including recent findings and how these tools are being used to develop strategies to improve cardiovascular health in European ancestry populations (which the authors acknowledge are the populations for which most data are available). This broader narrative and contextualization would more comprehensively describe the current state of this field, and help the reader understand how -omics approaches may be applied in South Asians. For instance, recent GWAS have been conducted in South Asian diaspora populations, and have contributed to a variety of strategies to understand cardiovascular risk in this population such as polygenic risk scoring (e.g., Wang M et al. “Validation of a genome-wide polygenic risk score for coronary artery disease in South Asians.” J Am Coll Cardiol. 2020;76(6):703-141.) This and other recent similar studies would be helpful to include.

We appreciate the comments and suggestions. In response to reviewer 1’s comments, we included a summary (Table 1) of recent omics studies of CVD in European ancestry and multi-ethnic populations. We also summarized growing methodologies and analytical tools which could be used to improve cardiovascular health among South Asians. We have added the following paragraph of “Strategies of omics studies” in the revised manuscript paper as below:

"Due to the rapidly growing technologies such as next generation sequencing and untargeted metabolomics, massive amount of accurate data can be obtained in a cost-effective manner. Therefore, there has been an urgent need for new methodologies and algorithms to form, for the purposes of computationally intensive data managing and analysis.[116] For GWAS, tools such as PLINK,[127] RVTESTS[128] and GENESIS[129] now allow the incorporation of many data pieces such as genotypes, annotations, allele frequencies and phenotypes to be processed efficiently in particular formats. For untargeted MWAS, mixOmics[130] and xmsPANDA (https://github.com/kuppal2/xmsPANDA) has been utilized to conduct feature selection, and the software Mummichog[131] for pathway exploration based on untargeted metabolomics data has been made available to bypass the feature identification stage. In addition, various tools have been developed for network analysis. For example, the weighted gene co-expression network analysis can be implemented using WGCNA[132] to identify clustering of genes, the software xMWAS[133]
can perform the integration and differential network analysis for multiple layers of omics data. The evolving omics researches of cardiovascular risk are coupled with these tools to achieve better understanding of the disease biological mechanism and pathway."

We agree that the polygenic risk score study by Wang M et al is a great example showing how to use existing GWAS findings and data of CAD to optimize the utility in South Asians. We added the following discussion about polygenic risk score in the “Genetic basis of atherosclerotic cardiovascular disease and heart failure” section as below:

"A large number of genetic associations have been identified in large population studies for CVD and risk factors. Although identified genetic associations have small effect sizes individually, polygenetic risk score (PRS) can combine such individual effects into a much stronger predictor of a disease trait. A few studies have shown the successful identifications of CAD and relevant traits.[61, 62] Earlier studies of PRS showed that the European ancestry-based can be transferred in other ancestry groups including South Asians, the associations of European ancestry-derived PRS were typically weaker in non-European ancestries. With large GWAS results from multiple ancestry groups, the PRS can be optimized to present ancestry-specific genetic risk for CVD. A study of 7,244 South Asian UK Biobank participants derived a PRS of CAD for South Asians from the previous GWAS findings that are primarily European-based. The PRS included 6,630,150 common variants, and demonstrated a successful framework for developing ancestry-specific PRS.[63] In another study, researchers identified significant association between the GRS, which comprised of 29 genome-wide significant blood pressure variants found among European descent, and blood pressure among South Asians.[54]"

2. Since the narrative review is cursory, it is not clear how the authors envision “multi-omic” methodologies should be deployed in South Asian populations. What are the key research questions with these methods in the South Asian population? What do the authors hypothesize will be found? Are there any characteristics specific to the South Asian population that should be accounted for (for example, emerging epidemiologic work identifies South Asian subgroups, e.g. Pakistani, Indian, etc.; how does this fit into research evaluating environmental influences on epigenomics or metabolomics)? Addressing these questions seems important for this review to understand how longitudinal multi-omics might “shed light” on CVD in South Asians.

We appreciate reviewer’s suggestion. We included a new section of “The multi-omics approach may address several challenges in CVD research facing South Asians” to discuss three main research questions as initial examples for future multi-omics research in South Asians as below:

"South Asians are an understudied population at high CVD risk even at low body weight and young ages, with a high propensity to diabetes, dyslipidemia, and hepatic steatosis, features that may provide new insights into CVD risk in the presence and absence of traditional risk factors. Given the genetic and molecular underpinning of CVD and risk factors, the genetic and molecular markers identified from multi-omics research may help accurately profile the CVD risk, given the unique characteristics among South Asians. Secondly, the multi-omics approach also holds the promise of revealing the heterogeneous mechanisms of CVD and risk factors. For example, HFpEF patients can be clustered into distinct subclasses with different clinical outcome using available demographic and clinical variables.[40, 41] T2DM is also a multi-factorial disease that involves numerous genetic
pathways and many environmental factors,[134] and has high prevalence in South Asian populations. Recent studies of T2MD subgroups have revealed striking heterogeneity of T2DM risk, etiology and outcomes, which may be further illustrated by multi-omic profiling. Lastly, multi-omics can help explain inter-individual variability in response to environmental risk exposures. Exposures to environmental risk such as air pollution (common and severer CVD risk in South Asia) are multifactorial and time-varying, thus, their measurements can be incomplete, costly, and imprecise for pathophysiological effects. Omic technologies such as epigenomics and metabolomics represent exogenous and endogenous effects related to CVD risk and have emerged as key components of exposome measures.[135, 136] They can capture the biological response to environmental exposures, thus, providing more precise risk assessment for subclinical and clinical CVD. Improved omics measurements of environmental exposure will also enable large scale gene-environment interaction study to further uncover molecular mechanisms underlying CVD pathophysiology."

"In addition to common genomic and multi-omics factors of CVD across ancestry groups, studying multi-omics among large population samples within South Asians may also discover genetic variants or molecular signatures unique for South Asians. Incorporating these multi-omic profiles can optimize the diagnosis, treatment and prognosis of CVD, which is the most important and still growing health burden for South Asian populations."

3. Are longitudinal -omics a component of research in other (including European ancestry) populations? What insights have been gained from such work? The authors suggest a longitudinal -omics approach in their concluding paragraphs, but have inadequately supported the potential merits of this strategy.

Unlike the genome profile, epigenomic, transcriptomic, proteomic and metabolomic profiles can be modified by environmental exposures, physiological conditions and disease status. The need for longitudinal multi-omics includes two components. First, current omics studies predominantly use a cross-sectional design, which enables the novel biomarker discovery but is limited to infer causal associations due to confounding and reverse causation. A longitudinal design including baseline omics and incident CVD would better demonstrate the prediction utility of omics markers for CVD progression, and has been implemented in omic association studies in other ancestry groups. Secondly, repeated measurement of omics profiles excluding genomics would be important to understand long-term risk and natural history of chronic diseases (PMID 22424236) such as CVD, since these omics layers can capture the varying environment and pathobiology related to CVD development. Such omics changes may also help identify reversible targets for novel interventions for CVD outcomes. However, the multi-omics study with repeated measures hasn't been reported for CVD. We added the following discussion to illustrate the need for longitudinal multi-omics in CVD research in the section of “The multi-omics approach is critical to understand CVD risk at the molecular level” as below:

"Unlike the genome profile, epigenomic, transcriptomic, proteomic and metabolomic profiles can be modified by environmental exposures, physiological conditions and disease status. Genomic data will complement epigenomic and metabolomic markers by identifying complex biological processes at the systems levels.[116] Current omics studies predominantly use a cross-sectional design, which enables the novel biomarker discovery but is limited to infer causal associations due to confounding and reverse causation. A longitudinal design including baseline omics and incident CVD would better demonstrate..."
the prediction utility of omics markers for CVD progression, and has been implemented in omic association studies in other ancestry groups. However, such a design doesn't incorporate the longitudinal changes of omics markers in relation to varying environment, pathobiology and disease progression. Repeated measurements of omics profiles excluding genomics would be important to understand long-term risk and natural history of chronic diseases[117] such as CVD. Such omics changes may also help identify reversible targets for novel interventions for CVD outcomes.

4. This review would be a prime opportunity to move this field forward beyond only stating that these -omics tools should be applied in South Asian populations. What are the existing challenges/barriers preventing these tools from being implemented in the study of South Asian CVD and CVD risk? (In parallel, why have most studies using -omics approaches occurred in European ancestry populations?) What strategies would help overcome these barriers? The authors are recommended to consider addressing these questions in their review to provide important context to the state of the science and a roadmap to achieve their recommended future directions.

We appreciate the suggestion. We revised the manuscript to address existing challenges and barriers for implementing multi-omics studies in South Asians in “Conclusion” section as below:

"Despite the available tools and fruitful research conducted among European ancestry, the implementation of similar research among South Asians remains challenging. Large high-quality studies are needed, with the requirements of high volume recruiting of population-based study samples, well-defined disease and phenotypes, long time follow-up, establishment of data registries, close collaborations between scientists with various expertise, such as study design, molecular biology and bioinformatics. There is a growing effort of research among the South Asian population, such as the Center for cArdio-metabolic Risk Reduction in South Asia (CARRS) study,[137] which is a large longitudinal study of 28,000 subjects across three large cities in South Asia. Studies that employ the integrative multi-omic approach to disentangling complex molecular systems underlying CVD are still anticipated."

5. The authors’ concluding paragraph introduces new concepts that were not previously discussed (for example, “HFpEF heterogeneity,” and the differentiation of South Asians in South Asia versus diaspora populations, which certainly would seem relevant with respect to environment-gene interaction). Consequently, their conclusions seem to be obscured and do not follow from their narrative. I recommended discussing these components of research gaps in South Asians earlier in the review, in order to justify their conclusions.

We included background of HFpEF heterogeneity and the role of gene-environment interaction in the “Introduction” section as below. We also discussed the research gaps in a new section in response to comment #2.

"Nearly half of the HF patients have HFpEF, with >90% being >60 years old, with rapidly increasing numbers.[32-35] Although numerous risk factors for HFpEF have been identified including hypertension, older age, female sex, obesity, diabetes, and renal disease,[36, 37] there are currently no class I guideline recommended treatments for HFpEF that improve mortality.[38, 39] HFpEF has received increased attention since HFpEF patients frequently experience delayed diagnosis and have limited treatment options. Recent studies have shown that HFpEF is a clinically heterogeneous disorder, consisting of subgroups with
related comorbidities and pathophysiologies, which lead to different progression trajectories.[40, 41] Employing latent class analytics and clustering techniques based on widely available clinical variables, several investigators have shown that patients with HFpEF can be divided into distinct classes with differing outcomes.[40, 41]"

"CVD progression can be affected by both gene and environment via different molecular pathways and mechanisms. Although the high throughput technologies have enabled accurate and cost-effective genotyping in large population samples, comparable high throughput measurements of environmental exposures are needed for large population studies. Gene-environment interaction is a common mechanism to explain complex disease risk, and inter-individual variability. Better understanding of gene-environment interactions and their causal relationships will point to pathways and mechanisms as potential targets for treatment and intervention. The gene-environment interaction study may also help understand the CVD risk among immigrant populations exposure to different environmental and lifestyle factors."

**Competing Interests:** No competing interests were disclosed.
3. If talking about multi-omics, the paper should also include publications of CVD and risk factors with proteomics and gene expression/RNA seq.

4. I also suggest the co-authors include a table summarize the resent publications for CVD outcomes and risk factors.

Is the topic of the review discussed comprehensively in the context of the current literature?
Partly

Are all factual statements correct and adequately supported by citations?
Yes

Is the review written in accessible language?
Yes

Are the conclusions drawn appropriate in the context of the current research literature?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical genetics and genomics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 07 May 2021
Yan Sun, Rollins School of Public Health, Emory University, Atlanta, USA

1. The title of this review is missing leading. When I read the title ‘Cardiovascular disease risk and pathophysiology in South Asians: can longitudinal multi-omics shed light’ I thought this review summaries omics research of CVD and risk factors in South Asia. But the paper mostly based on participants of European origin. Therefore, this is really a perspective of future research in Asia. The title is not appropriate.

The reviewer is correct about the perspective component of the manuscript because the multi-omics research of CVD is still limited, particularly among South Asians. Meanwhile, the prevalence of heart failure and atherosclerotic heart disease in South Asian populations is rising, and the incidence is higher at younger age and leaner body mass index in South Asians compared to European population. The growing evidence from epidemiologic studies emphasizes the importance of improving cardiovascular health among South Asians. Therefore, we acknowledge that the current omics research of CVD is primarily conducted among European population, and address the needs for such omics studies in South Asians. We also noted a recent genetic study of coronary heart disease, which demonstrated the predictive ability of polygenic risk score in South Asians (PMID 32762905). As a result, we hope the title
represents the current knowledge of CVD in South Asians, multi-omics of CVD (mostly in European ancestry) and anticipation for the future efforts in tackling the disparity in South Asians, a mixture of review and perspective in the present manuscript.

2. I suggest the paper comprehensively summarized the research publications in the field. For example, for blood pressure GWAS, the paper cited 2011 GWAS. There are at least 4 GWAS publications after 2015. For DNA methylation, there are many papers with larger sample sizes published after 2015. But this paper only included a few for CVD risk factors.

We appreciate the reviewer's suggestion on including recent publications of large GWAS and DNA methylation studies. We included recent blood pressure GWAS with larger sample sizes (PMID 30224653, PMID 30578418, PMID 30429575), in addition to the 2011 blood pressure GWAS as an early multi-ethnic GWAS including South Asians (PMID 21909115). We also added recent multi-ethnic GWAS publications of CVD and risk factors, including coronary heart disease (PMID 33532862, PMID 33020668, 2021 preprint DOI: 10.21203/rs.3.rs-275591/v1), heart failure (PMID 31919418, PMID 30586722), type 2 diabetes (PMID 32541925), BMI (PMID 30124842), and blood lipids (PMID 30275531). The revised text under “Genetic basis of atherosclerotic cardiovascular disease and heart failure” is listed below:

"Genome-wide association studies (GWAS) have identified a large number of genetic loci associated with coronary heart disease,[42-44] ASCVD,[45-47] HF,[48, 49] and their risk factors, such as BMI,[50, 51] blood lipids,[52, 53] blood pressure,[54-57] and T2DM;[58] however, these genetic variations explain only a small portion of risk in populations.[59, 60]"

When reviewing the epigenome-wide association studies (EWAS), we focused on studies with prospective design and potential causal inference. For example, the Agha 2019 study (PMID 31424985) discovered CpG sites associated with incident coronary heart disease or myocardial infarction, and the Mendelian Randomization analysis revealed the causal effect of DNA methylation sites in the disease mechanism. The Westerman's 2019 paper (PMID 31615550) also discovered DNA methylation modules associated with incident cardiovascular disease. Additionally, we included several large EWAS of CVD risk factors, including type 2 diabetes (PMID 31506343, PMID 26095709), BMI (PMID 28002404), blood lipids (PMID 28194238) and blood pressure (PMID 31999706, PMID 32520614, PMID 29198723). The revised text under “Epigenetic signatures identify molecular pathways of CVD that are activated in the context of environmental risk” is listed below:

"EWAS provides an unbiased approach for identifying molecular mediators of genetic and environmental factors that may explain residual risk of disease[74] and has been applied to investigate CVD[75-79] and risk factors.[80-85]"

"The study of incident T2DM conducted among Indian Asians and Europeans showed a 2.5 times higher adjusted risk among Indian Asians than Europeans, and five loci including ABCG1, PHOSPHO1, SOCS3, SREBF1, and TXNIP were associated with incident T2DM among Indian Asians and replicated among Europeans.[81] The other study of incident T2DM reported additional loci such as PHGDH and CPT1A, which were discovered among Europeans and replicated among Indian Asians.[80] The study of blood pressure compared the methylation profiles between Europeans and South Asians revealed many distinct loci between the two ancestries with a small overlap.[84] To our current knowledge, some known trans-ethnic epigenetic loci might be transportable from Europeans to South Asians,
but there is still a lack in studies specifically among large South Asian populations to further explore novel loci to explain the higher risk. A number of EWAS have examined DNAm patterns associated with air pollution, which is a known environmental risk factor for CVD.[97,98]"

For the multi-omics studies, we included a recent publication that integrated genome-wide single-nucleotide polymorphisms, gene expression, and DNA methylation, revealed the molecular mechanism of heart failure (PMID 31703168), and another study of CVD (PMID 33836805). The revised text under “The multi-omics approach is critical to understand CVD risk at the molecular level” is listed below:

"A recent study using longitudinal big data including genome, immunome, transcriptome, proteome, metabolome, microbiome and wearable monitoring have showed the potential of revealing cardiovascular pathophysiology on a molecular basis.[118] Although increasing number of population studies have measured multi-omics data, recent studies have focused on single -omic association study and used additional -omics data to better understand the molecular functions of identified associations.[119] Integrated multi-omics studies of CVD outcomes have been limited. A large HF study of over 8,000 participants within the Framingham Heart Study utilized integrative trans-omics data including genetic variations, DNA methylation, and gene expression data to reveal genetic contributions towards HF. The transportability of such findings to South Asians needs to be evaluated.[120] Another study integrating DNA methylation and gene expression data identified independent latent factors associated with CVD. The unsupervised machine learning of multi-omics successfully improved classification and discrimination.[121]"

3. If talking about multi-omics, the paper should also include publications of CVD and risk factors with proteomics and gene expression/RNA seq.

We appreciate the suggestion. We have added a paragraph summarizing “Proteomics and gene expression reveal the interactions between cell processes and external environment". The text under this section is listed below:

"The proteomics technologies have evolved rapidly in the past two decades. Recent applications of proteomics in population studies have produced interesting findings in CVD research.[99] A recent study of the healthy human heart tissue collected from autopsy determined the healthy heart proteome. The resulted database, which included over 10,700 proteins, is a comprehensive resource for the downstream investigations.[100] In a targeted proteomics study, two protein signatures for high-risk plaques and absence of coronary atherosclerosis were identified among a cohort with suspected coronary artery disease.[101] The prediction accuracy of a model constructed by 50 proteins showed a better prediction accuracy in adverse cardiovascular events than a traditional risk factor model.[102] These findings are critical in CVD risk prediction and differentiation of CVD subtypes. In addition, gene expression profiling in CVD also enables a better understanding of pathophysiology of CVD.[103] Patterns of gene expression of human aorta tissue was investigated to identify genes with prediction power in atherosclerosis.[104] Studies of non-coding transcriptome, which even though have limited protein-coding functions, are discovering the pathology of CVD. To eventually realize precision medicine, we are still facing the challenges of standardization of methodologies and translation.[105] In addition, researches in this area largely depend on the availability of tissues from autopsy, thus the clinical translation and implementation has been limited. However, such studies facilitate a
better understanding on the disease causal pathway and mechanism. Particularly, there has been a lack of similar researches particularly among South Asians, which is urging the future efforts of investigation."

4. I also suggest the co-authors include a table summarize the resent publications for CVD outcomes and risk factors.

We appreciate the suggestion. As we focused on the CVD outcomes in South Asians, we added a summary table (Table 1) for recent omics studies of CVD including coronary artery disease, stroke, myocardial infarction and heart failure, particular in multi-ethnic studies when available. We also added text in the “Introduction” section.

Competing Interests: No competing interests were disclosed.