REVIEW

REVISED Cardiovascular disease risk and pathophysiology in South Asians: can longitudinal multi-omics shed light? [version 2; peer review: 2 approved]

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

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

v2 **First published:** 27 Oct 2020, **5**:255
https://doi.org/10.12688/wellcomeopenres.16336.1
Latest published: 20 May 2021, **5**:255
https://doi.org/10.12688/wellcomeopenres.16336.2

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 (HFpEF) and reduced ejection fraction (HFrEF). The prevalence of HF in South Asians is also rising with other metabolic diseases, and HFpEF 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.

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
version 2 (revision) 20 May 2021	✓ report	✓ report
version 1 27 Oct 2020	? report	? report

1. **Chunyu Liu**, Boston University, Boston, USA

2. **Nilay Shah**, Northwestern University Feinberg School of Medicine, Chicago, USA

Any reports and responses or comments on the

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

article can be found at the end of the article.

This article is included in the Wellcome Trust/DBT India Alliance gateway.

Corresponding author: Yan V. Sun (yan.v.sun@emory.edu)

Author roles: Sun YV: Conceptualization, Data Curation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Liu C**: Data Curation, Writing – Original Draft Preparation, Writing – Review & Editing; **Staimez L**: Writing – Review & Editing; **Ali MK**: Writing – Review & Editing; **Chang H**: Writing – Review & Editing; **Kondal D**: Writing – Review & Editing; **Patel S**: Writing – Review & Editing; **Jones D**: Writing – Review & Editing; **Mohan V**: Writing – Review & Editing; **Tandon N**: Writing – Review & Editing; **Prabhakaran D**: Writing – Review & Editing; **Quyyumi AA**: Conceptualization, Writing – Review & Editing; **Narayan KMV**: Conceptualization, Writing – Review & Editing; **Agrawal A**: Methodology, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Dr. Anurag Agrawal is supported by The Wellcome Trust DBT India Alliance (grant IA/CPHS/14/1/501489). The CARRS Study was funded in part by the National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Department of Health and Human Services, under Contract No. HHSN268200900026C, and the United Health Group, Minneapolis, MN, USA. K M Venkat Narayan (Narayan), Mohammed K Ali (MKA), Unjali P. Gujral (UPG), Shivani A. Patel (SAP) were funded in part by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Number P30DK111024. Narayan was funded in part for "Worksite Lifestyle Program for Reducing Diabetes and Cardiovascular Risk in India" project funded by NHLBI, NIH, Department of Health and Human Services under Award number R01HL125442. SAP, Narayan, MKA, Nikhil Tandon (NT), Dorairaj Prabhakaran (DP) were supported in part by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH), award number 5U01HL138635 under the Hypertension Outcomes for T4 Research within Lower Middle-Income Countries (Hy-TREC) program. Dimple Kondal (DK) has been supported by Fogarty International Center for PH leader Course, National Institutes of Health under grant number D43TW009135.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Sun YV, Liu C, Staimez L *et al.* **Cardiovascular disease risk and pathophysiology in South Asians: can longitudinal multi-omics shed light? [version 2; peer review: 2 approved]** Wellcome Open Research 2021, **5**:255 https://doi.org/10.12688/wellcomeopenres.16336.2

First published: 27 Oct 2020, **5**:255 https://doi.org/10.12688/wellcomeopenres.16336.1

REVISED **Amendments from Version 1**

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[11–18]. 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[19–27].

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 [HFpEF] 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[28]. The prevalence of HF in South Asians is also rising with other metabolic diseases[29,30], and HFpEF develops at younger age and leaner BMI than for Whites[31].

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.

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[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]. 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, polygenetic risk score (PRS) can combine such individual effects into a much stronger predicter 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].

Current findings in CVD genetics are overwhelmingly driven by European ancestry, that disproportionally represents the majority of the global population at risk[64]. Recent studies start

Table 1. Summary of selected omics studies of CVD.

PMID	First Author	Year	Phenotype	Sample size	Ethnicity
Genetic basis of atherosclerotic cardiovascular disease and heart failure					
33532862	Hartiala JA	2021	Myocardial infarction	Discovery ~61,000 cases, 577,000 controls Replication: ~165,000+27,000 subjects	European, East Asian
33020668	Koyama S	2020	CAD	121,234 cases, 527,824 controls	East Asian, European
N/A DOI:10.21203/rs.3.rs-275591/v1	Assimes T	2021	CAD	~250,000 cases	European, African American, Hispanic, East Asian
26343387	Nikpay M	2015	CAD	60,801 cases, 123,504 controls	European, South Asian, East Asian, Hispanic, African American
29212778	van der Harst P	2018	CAD	Discovery: 34,541 cases, 261,984 controls Replication: 88,192 cases, 162,544 controls	European
31160810	Malik R	2018	Stroke	67,162 cases, 454,450 controls	European, East Asian, African, South Asian, mixed Asian, Latin American
31919418	Shah S	2020	HF	47,309 cases, 930,014 controls	European
23202125	Consortium CAD	2013	CAD	63,746 cases, 130,681 controls	European, South Asian
Epigenetic signatures identify molecular pathways of CVD that are activated in the context of environmental risk					
31424985	Agha G	2019	Myocardial infarction, CAD	11,461 subjects	European, African American
28515798	Nakatochi M	2017	Myocardial infarction	192 cases, 192 controls	Japanese
28172975	Rask-Andersen M	2016	Myocardial infarction	729 subjects	European
31615550	Westerman K	2019	CVD	Discovery: 2,023 subjects Validation: 2,587 subjects	European, African American, Hispanic
28838933	Meder B	2017	HF	Discovery: 41 cases, 31+31 controls Replication 1: 18+9 cases, 8+28 controls Replication 2: 82 cases, 109 controls	European
Proteomics and gene expression reveal the interactions between cell processes and external environment					
30587458	Bom MJ	2019	Coronary plaque morphology	196 subjects	European
32808014	Hoogeveen RM	2020	CVD	Discovery: 822 subjects Validation: 702 subjects	European

Wellcome Open Research 2021, 5:255 Last updated: 03 JUN 2021

PMID	First Author	Year	Phenotype	Sample size	Ethnicity
Metabolomic signatures present complex metabolic state and environmental exposure					
25881932	Cheng ML	2015	HF	Discovery: 183 cases, 51 controls Validation: 218 cases, 63 controls	East Asian
31092011	Wang Z	2019	CAD	3,598 subjects	African American, European
29893901	Bhupathiraju SN	2018	Cardiometabolic risk	145	South Asian
23788672	Zheng Y	2013	HF	1,744 subjects	African American
29096792	Lanfear DE	2017	HF	Discovery: 516 subjects Validation: 516 subjects	European, African American
The multi-omics approach is critical to understand CVD risk at the molecular level					
29096792	Andersson C	2019	HF	8,372 subjects	European
33836805	Palou-Marquez G	2021	CVD	Methylation: 2,055 subjects Gene expression: 914 subjects	European

showing the benefits of including diverse ancestries in discovering novel genetic loci associated with diseases[65], 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[66]. Epigenetic modification, through DNA methylation (DNAm) and other molecular mechanisms can regulate gene expression levels that can influence susceptibility to disease development[67], including atherosclerosis[68] and chronic inflammation[69], two important pathophysiological processes leading to CVD. Furthermore, epigenetic markers are modified by age[70,71] and environmental risk factors, such as smoking[72] and poor nutrition[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[74]. 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].

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[75]. 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[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[70,86,87], BMI[88,89], diabetes[80,81,90,91], blood lipids[92], smoking[72,93,94] and inflammation[95]. These DNAm markers may provide a measure of longitudinal CVD risk[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[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-ethic 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]. 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[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.

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[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[109]. 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[110]. 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 that 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[111]. Recent studies also reported changes in global metabolism in relation to HF risk[112] and outcomes[106]. 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[106,113]. 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[74,114,115]. Recent GWAS have identified >160 genomic loci for CAD[46], 11 loci for HF[49], and >300 loci for T2DM[58]. 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[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 revers 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.

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]. Additionally, our recent joint epigenomics-metabolomics study of smoking[122] 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[123–125]. Longitudinal design can also distinguish cause from consequence using appropriate modeling[126]. 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[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 understating 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 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.

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 cArdiometabolic 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. 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

1. Volgman AS, Palaniappan LS, Aggarwal NT, *et al.*: **Atherosclerotic Cardiovascular Disease in South Asians in the United States: Epidemiology, Risk Factors, and Treatments: A Scientific Statement From the American Heart Association.** *Circulation.* 2018; **138**(1): e1–e34.
PubMed Abstract | Publisher Full Text
2. Prabhakaran D, Jeemon P, Roy A: **Cardiovascular Diseases in India: Current Epidemiology and Future Directions.** *Circulation.* 2016; **133**(16): 1605–20.
PubMed Abstract | Publisher Full Text
3. Gupta R, Guptha S, Sharma KK, *et al.*: **Regional variations in cardiovascular risk factors in India: India heart watch.** *World J Cardiol.* 2012; **4**(4): 112–20.
PubMed Abstract | Publisher Full Text | Free Full Text
4. **IDF Diabetes Atlas**. 6th ed. 2013.
Reference Source
5. Gupta R, Guptha S, Agrawal A, *et al.*: **Secular trends in cholesterol lipoproteins and triglycerides and prevalence of dyslipidemias in an urban Indian population.** *Lipids Health Dis.* 2008; **7**: 40.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Anchala R, Kannuri NK, Pant H, *et al.*: **Hypertension in India: a systematic review and meta-analysis of prevalence, awareness, and control of hypertension.** *J Hypertens.* 2014; **32**(6): 1170–7.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Kearney PM, Whelton M, Reynolds K, *et al.*: **Global burden of hypertension: analysis of worldwide data.** *Lancet.* 2005; **365**(9455): 217–23.
PubMed Abstract | Publisher Full Text
8. Fernandez-Friera L, Fuster V, Lopez-Melgar B, *et al.*: **Normal LDL-Cholesterol Levels Are Associated With Subclinical Atherosclerosis in the Absence of Risk Factors.** *J Am Coll Cardiol.* 2017; **70**(24): 2979–91.
PubMed Abstract | Publisher Full Text
9. Lee K: **Muscle Mass and Body Fat in Relation to Cardiovascular Risk Estimation and Lipid-Lowering Eligibility.** *J Clin Densitom.* 2017; **20**(2): 247–55.
PubMed Abstract | Publisher Full Text
10. Nambi V, Bhatt DL: **Primary Prevention of Atherosclerosis: Time to Take a Selfie?** *J Am Coll Cardiol.* 2017; **70**(24): 2992–4.
PubMed Abstract | Publisher Full Text
11. Levy D, Garrison RJ, Savage DD, *et al.*: **Left ventricular mass and incidence of coronary heart disease in an elderly cohort. The Framingham Heart Study.** *Ann Intern Med.* 1989; **110**(2): 101–7.
PubMed Abstract | Publisher Full Text
12. Salonen JT, Salonen R: **Ultrasonographically assessed carotid morphology and the risk of coronary heart disease.** *Arterioscler Thromb.* 1991; **11**(5): 1245–9.
PubMed Abstract | Publisher Full Text
13. O'Leary DH, Polak JF, Kronmal RA, *et al.*: **Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. Cardiovascular Health Study Collaborative Research Group.** *N Engl J Med.* 1999; **340**(1): 14–22.
PubMed Abstract | Publisher Full Text
14. Heald CL, Fowkes FGR, Murray GD, *et al.*: **Risk of mortality and cardiovascular disease associated with the ankle-brachial index: Systematic review.** *Atherosclerosis.* 2006; **189**(1): 61–9.
PubMed Abstract | Publisher Full Text
15. Criqui MH, McClelland RL, McDermott MM, *et al.*: **The ankle-brachial index and incident cardiovascular events in the MESA (Multi-Ethnic Study of Atherosclerosis).** *J Am Coll Cardiol.* 2010; **56**(18): 1506–12.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Budoff MJ, Young R, Burke G, *et al.*: **Ten-year association of coronary artery calcium with atherosclerotic cardiovascular disease (ASCVD) events: the multi-ethnic study of atherosclerosis (MESA).** *Eur Heart J.* 2018; **39**(25): 2401–8.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Greenland P, Blaha MJ, Budoff MJ, *et al.*: **Coronary Calcium Score and Cardiovascular Risk.** *J Am Coll Cardiol.* 2018; **72**(4): 434–47.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Mitchell CC, Korcarz CE, Gepner AD, *et al.*: **Carotid Artery Echolucency, Texture Features, and Incident Cardiovascular Disease Events: The MESA Study.** *J Am Heart Assoc.* 2019; **8**(3): e010875.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Pais P, Pogue J, Gerstein H, *et al.*: **Risk factors for acute myocardial infarction in Indians: a case-control study.** *Lancet.* 1996; **348**(9024): 358–63.
PubMed Abstract | Publisher Full Text
20. Gupta M, Doobay AV, Singh N, *et al.*: **Risk factors, hospital management and outcomes after acute myocardial infarction in South Asian Canadians and matched control subjects.** *CMAJ.* 2002; **166**(6): 717–22.
PubMed Abstract | Free Full Text
21. Singh N, Gupta M: **Clinical characteristics of South Asian patients hospitalized with heart failure.** *Ethn Dis.* 2005; **15**(4): 615–9.
PubMed Abstract
22. Joshi P, Islam S, Pais P, *et al.*: **Risk factors for early myocardial infarction in South Asians compared with individuals in other countries.** *JAMA.* 2007; **297**(3): 286–94.
PubMed Abstract | Publisher Full Text
23. Xavier D, Pais P, Devereaux PJ, *et al.*: **Treatment and outcomes of acute coronary syndromes in India (CREATE): a prospective analysis of registry data.** *Lancet.* 2008; **371**(9622): 1435–42.
PubMed Abstract | Publisher Full Text
24. Chen Y, Copeland WK, Vedanthan R, *et al.*: **Association between body mass index and cardiovascular disease mortality in east Asians and south Asians: pooled analysis of prospective data from the Asia Cohort Consortium.** *BMJ.* 2013; **347**: f5446.
PubMed Abstract | Publisher Full Text | Free Full Text
25. Anjana RM, Pradeepa R, Das AK, *et al.*: **Physical activity and inactivity patterns in India - results from the ICMR-INDIAB study (Phase-1) [ICMR-INDIAB-5].** *Int J Behav Nutr Phys Act.* 2014; **11**(1): 26.
PubMed Abstract | Publisher Full Text | Free Full Text
26. Huffman M, Jeemon P, Prabhakaran D, *et al.*: **A race against time II: the challenge of cardiovascular diseases in developing economies**. 2014.
Reference Source
27. Harikrishnan S, Leeder S, Huffman M, *et al.*: **A Race against Time: The Challenge of Cardiovascular Disease in Developing Economies**. 2nd ed: Centre for Chronic Disease Control; 2014.
Reference Source
28. Heidenreich PA, Albert NM, Allen LA, *et al.*: **Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association.** *Circ Heart Fail.* 2013; **6**(3): 606–19.
PubMed Abstract | Publisher Full Text | Free Full Text
29. Huffman MD, Prabhakaran D: **Heart failure: epidemiology and prevention in India.** *Natl Med J India.* 2010; **23**(5): 283–8.
PubMed Abstract | Free Full Text
30. Guha S, Harikrishnan S, Ray S, *et al.*: **CSI position statement on management of heart failure in India.** *Indian Heart J.* 2018; **70 Suppl 1**(Suppl 1): S1–S72.
PubMed Abstract | Publisher Full Text | Free Full Text
31. Tromp J, Teng TH, Tay WT, *et al.*: **Heart failure with preserved ejection fraction in Asia.** *Eur J Heart Fail.* 2019; **21**(1): 23–36.
PubMed Abstract | Publisher Full Text
32. Bhatia RS, Tu JV, Lee DS, *et al.*: **Outcome of heart failure with preserved ejection fraction in a population-based study.** *N Engl J Med.* 2006; **355**(3): 260–9.
PubMed Abstract | Publisher Full Text
33. Owan TE, Hodge DO, Herges RM, *et al.*: **Trends in prevalence and outcome of heart failure with preserved ejection fraction.** *N Engl J Med.* 2006; **355**(3): 251–9.
PubMed Abstract | Publisher Full Text
34. Lee DS, Gona P, Vasan RS, *et al.*: **Relation of disease pathogenesis and risk factors to heart failure with preserved or reduced ejection fraction: insights from the framingham heart study of the national heart, lung, and blood institute.** *Circulation.* 2009; **119**(24): 3070–7.
PubMed Abstract | Publisher Full Text | Free Full Text
35. Borlaug BA, Paulus WJ: **Heart failure with preserved ejection fraction: pathophysiology, diagnosis, and treatment.** *Eur Heart J.* 2011; **32**(6): 670–9.
PubMed Abstract | Publisher Full Text | Free Full Text
36. Yancy CW, Jessup M, Bozkurt B, *et al.*: **2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines.** *J Am Coll Cardiol.* 2013; **62**(16): e147–239.
PubMed Abstract | Publisher Full Text
37. Andersson C, Vasan RS: **Epidemiology of heart failure with preserved ejection fraction.** *Heart Fail Clin.* 2014; **10**(3): 377–88.
PubMed Abstract | Publisher Full Text | Free Full Text
38. Udelson JE: **Heart failure with preserved ejection fraction.** *Circulation.* 2011; **124**(21): e540–3.
PubMed Abstract | Publisher Full Text
39. Brouwers FP, de Boer RA, van der Harst P, *et al.*: **Incidence and epidemiology of new onset heart failure with preserved vs. reduced ejection fraction in a community-based cohort: 11-year follow-up of PREVEND.** *Eur Heart J.* 2013; **34**(19): 1424–31.
PubMed Abstract | Publisher Full Text
40. Kao DP, Lewsey JD, Anand IS, *et al.*: **Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response.** *Eur J Heart Fail.* 2015; **17**(9): 925–35.
PubMed Abstract | Publisher Full Text | Free Full Text
41. Shah SJ, Katz DH, Selvaraj S, *et al.*: **Phenomapping for novel classification of heart failure with preserved ejection fraction.** *Circulation.* 2015; **131**(3): 269–79.
PubMed Abstract | Publisher Full Text | Free Full Text
42. Hartiala JA, Han Y, Jia Q, *et al.*: **Genome-wide analysis identifies novel susceptibility loci for myocardial infarction.** *Eur Heart J.* 2021; **42**(9): 919–33.
PubMed Abstract | Publisher Full Text | Free Full Text
43. Koyama S, Ito K, Terao C, *et al.*: **Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease.** *Nat Genet.* 2020; **52**(11): 1169–77.
PubMed Abstract | Publisher Full Text

44. Assimes T, Tcheandjieu C, Zhu X: **A large-scale multi-ethnic genome-wide association study of coronary artery disease**. PREPRINT (Version 2) available at Research Square.
Publisher Full Text

45. Nikpay M, Goel A, Won HH, *et al.*: **A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease.** *Nat Genet.* 2015; **47**(10): 1121–30.
PubMed Abstract | Publisher Full Text | Free Full Text

46. van der Harst P, Verweij N: **Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease.** *Circ Res.* 2018; **122**(3): 433–43.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Malik R, Chauhan G, Traylor M, *et al.*: **Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes.** *Nat Genet.* 2018; **50**(4): 524–37.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Aragam KG, Chaffin M, Levinson RT, *et al.*: **Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery.** *Circulation.* 2018.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Shah S, Henry A, Roselli C, *et al.*: **Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure.** *Nat Commun.* 2020; **11**(1): 163.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Speliotes EK, Willer CJ, Berndt SI, *et al.*: **Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.** *Nat Genet.* 2010; **42**(11): 937–48.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Yengo L, Sidorenko J, Kemper KE, *et al.*: **Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry.** *Hum Mol Genet.* 2018; **27**(20): 3641–9.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Teslovich TM, Musunuru K, Smith AV, *et al.*: **Biological, clinical and population relevance of 95 loci for blood lipids.** *Nature.* 2010; **466**(7307): 707–13.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Klarin D, Damrauer SM, Cho K, *et al.*: **Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program.** *Nat Genet.* 2018; **50**(11): 1514–23.
PubMed Abstract | Publisher Full Text | Free Full Text

54. International Consortium for Blood Pressure Genome-Wide Association Studies; Ehret GB, Munroe PB, *et al.*: **Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk.** *Nature.* 2011; **478**(7367): 103–9.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Giri A, Hellwege JN, Keaton JM, *et al.*: **Trans-ethnic association study of blood pressure determinants in over 750,000 individuals.** *Nat Genet.* 2019; **51**(1): 51–62.
PubMed Abstract | Publisher Full Text | Free Full Text

56. Evangelou E, Warren HR, Mosen-Ansorena D, *et al.*: **Publisher Correction: Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits.** *Nat Genet.* 2018; **50**(12): 1755.
PubMed Abstract | Publisher Full Text

57. Evangelou E, Warren HR, Mosen-Ansorena D, *et al.*: **Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits.** *Nat Genet.* 2018; **50**(10): 1412–25.
PubMed Abstract | Publisher Full Text | Free Full Text

58. Vujkovic M, Keaton JM, Lynch JA, *et al.*: **Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis.** *Nat Genet.* 2020; **52**(7): 680–91.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Nikpay M, Goel A, Won HH, *et al.*: **A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease.** *Nat Genet.* 2015; **47**(10): 1121–30.
PubMed Abstract | Publisher Full Text | Free Full Text

60. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, *et al.*: **Large-scale association analysis identifies new risk loci for coronary artery disease.** *Nat Genet.* 2013; **45**(1): 25–33.
PubMed Abstract | Publisher Full Text | Free Full Text

61. Hindy G, Aragam KG, Ng K, *et al.*: **Genome-Wide Polygenic Score, Clinical Risk Factors, and Long-Term Trajectories of Coronary Artery Disease.** *Arterioscler Thromb Vasc Biol.* 2020; **40**(11): 2738–46.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Kathiresan S, Melander O, Anevski D, *et al.*: **Polymorphisms associated with cholesterol and risk of cardiovascular events.** *N Engl J Med.* 2008; **358**(12): 1240–9.
PubMed Abstract | Publisher Full Text

63. Wang M, Menon R, Mishra S, *et al.*: **Validation of a Genome-Wide Polygenic Score for Coronary Artery Disease in South Asians.** *J Am Coll Cardiol.* 2020; **76**(6): 703–14.
PubMed Abstract | Publisher Full Text | Free Full Text

64. Gurdasani D, Barroso I, Zeggini E, *et al.*: **Genomics of disease risk in globally diverse populations.** *Nat Rev Genet.* 2019; **20**(9): 520–35.
PubMed Abstract | Publisher Full Text

65. Wojcik GL, Graff M, Nishimura KK, *et al.*: **Genetic analyses of diverse populations improves discovery for complex traits.** *Nature.* 2019; **570**(7762): 514–8.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Bird A: **Perceptions of epigenetics.** *Nature.* 2007; **447**(7143): 396–8.
PubMed Abstract | Publisher Full Text

67. Ordovas JM, Smith CE: **Epigenetics and cardiovascular disease.** *Nat Rev Cardiol.* 2010; **7**(9): 510–9.
PubMed Abstract | Publisher Full Text | Free Full Text

68. Turunen MP, Aavik E, Yla-Herttuala S: **Epigenetics and atherosclerosis.** *Biochim Biophys Acta.* 2009; **1790**(9): 886–91.
PubMed Abstract | Publisher Full Text

69. Stenvinkel P, Karimi M, Johansson S, *et al.*: **Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease?** *J Intern Med.* 2007; **261**(5): 488–99.
PubMed Abstract | Publisher Full Text

70. Teschendorff AE, Menon U, Gentry-Maharaj A, *et al.*: **Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.** *Genome Res.* 2010; **20**(4): 440–6.
PubMed Abstract | Publisher Full Text | Free Full Text

71. Bocklandt S, Lin W, Sehl ME, *et al.*: **Epigenetic predictor of age.** *PLoS One.* 2011; **6**(6): e14821.
PubMed Abstract | Publisher Full Text | Free Full Text

72. Breitling LP, Yang R, Korn B, *et al.*: **Tobacco-smoking-related differential DNA methylation: 27K discovery and replication.** *Am J Hum Genet.* 2011; **88**(4): 450–7.
PubMed Abstract | Publisher Full Text | Free Full Text

73. Waterland RA, Jirtle RL: **Transposable elements: targets for early nutritional effects on epigenetic gene regulation.** *Mol Cell Biol.* 2003; **23**(15): 5293–300.
PubMed Abstract | Publisher Full Text | Free Full Text

74. Sun YV: **The Influences of Genetic and Environmental Factors on Methylome-wide Association Studies for Human Diseases.** *Curr Genet Med Rep.* 2014; **2**(4): 261–70.
PubMed Abstract | Publisher Full Text | Free Full Text

75. Agha G, Mendelson MM, Ward-Caviness CK, *et al.*: **Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease.** *Circulation.* 2019; **140**(8): 645–57.
PubMed Abstract | Publisher Full Text | Free Full Text

76. Nakatochi M, Ichihara S, Yamamoto K, *et al.*: **Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease.** *Clin Epigenetics.* 2017; **9**: 54.
PubMed Abstract | Publisher Full Text | Free Full Text

77. Rask-Andersen M, Martinsson D, Ahsan M, *et al.*: **Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction.** *Hum Mol Genet.* 2016; **25**(21): 4739–48.
PubMed Abstract | Publisher Full Text

78. Westerman K, Sebastiani P, Jacques P, *et al.*: **DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure.** *Clin Epigenetics.* 2019; **11**: 142.
PubMed Abstract | Publisher Full Text | Free Full Text

79. Meder B, Haas J, Sedaghat-Hamedani F, *et al.*: **Epigenome-Wide Association Study Identifies Cardiac Gene Patterning and a Novel Class of Biomarkers for Heart Failure.** *Circulation.* 2017; **136**(16): 1528–44.
PubMed Abstract | Publisher Full Text

80. Cardona A, Day FR, Perry JRB, *et al.*: **Epigenome-Wide Association Study of Incident Type 2 Diabetes in a British Population: EPIC-Norfolk Study.** *Diabetes.* 2019; **68**(12): 2315–26.
PubMed Abstract | Publisher Full Text | Free Full Text

81. Chambers JC, Loh M, Lehne B, *et al.*: **Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study.** *Lancet Diabetes Endocrinol.* 2015; **3**(7): 526–34.
PubMed Abstract | Publisher Full Text | Free Full Text

82. Wahl S, Drong A, Lehne B, *et al.*: **Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.** *Nature.* 2017; **541**(7635): 81–6.
PubMed Abstract | Publisher Full Text | Free Full Text

83. Braun KVE, Dhana K, de Vries PS, *et al.*: **Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study.** *Clin Epigenetics.* 2017; **9**: 15.
PubMed Abstract | Publisher Full Text | Free Full Text

84. Kazmi N, Elliott HR, Burrows K, *et al.*: **Associations between high blood pressure and DNA methylation.** *PLoS One.* 2020; **15**(1): e0227728.
PubMed Abstract | Publisher Full Text | Free Full Text

85. Richard MA, Huan T, Ligthart S, *et al.*: **DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation.** *Am J Hum Genet.* 2017; **101**(6): 888–902.
PubMed Abstract | Publisher Full Text | Free Full Text

86. Hannum G, Guinney J, Zhao L, *et al.*: **Genome-wide methylation profiles reveal quantitative views of human aging rates.** *Mol cell.* 2013; **49**(2): 359–67.
PubMed Abstract | Publisher Full Text | Free Full Text

87. Horvath S: **DNA methylation age of human tissues and cell types.** *Genome Biol.* 2013; **14**(10): R115.
PubMed Abstract | Publisher Full Text | Free Full Text

88. Dick KJ, Nelson CP, Tsaprouni L, *et al.*: **DNA methylation and body-mass index:**

a genome-wide analysis. *Lancet.* 2014; **383**(9933): 1990–8.
PubMed Abstract | Publisher Full Text

89. Demerath EW, Guan W, Grove ML, *et al.*: **Epigenome-wide association study (EWAS) of BMI BMI change and waist circumference in African American adults identifies multiple replicated loci.** *Hum Mol Genet.* 2015; **24**(15): 4464–79.
PubMed Abstract | Publisher Full Text | Free Full Text

90. Meeks KAC, Henneman P, Venema A, *et al.*: **Epigenome-wide association study in whole blood on type 2 diabetes among sub-Saharan African individuals: findings from the RODAM study.** *Int J Epidemiol.* 2019; **48**(1): 58–70.
PubMed Abstract | Publisher Full Text | Free Full Text

91. Mathur R, Hui Q, Huang Y, *et al.*: **DNA Methylation Markers of Type 2 Diabetes Mellitus Among Male Veterans With or Without Human Immunodeficiency Virus Infection.** *J Infect Dis.* 2019; **219**(12): 1959–62.
PubMed Abstract | Publisher Full Text | Free Full Text

92. Irvin MR, Zhi D, Joehanes R, *et al.*: **Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study.** *Circulation.* 2014; **130**(7): 565–72.
PubMed Abstract | Publisher Full Text | Free Full Text

93. Shenker NS, Polidoro S, van Veldhoven K, *et al.*: **Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking.** *Hum Mol Genet.* 2013; **22**(5): 843–51.
PubMed Abstract | Publisher Full Text

94. Sun YV, Smith AK, Conneely KN, *et al.*: **Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans.** *Hum genet.* 2013; **132**(9): 1027–37.
PubMed Abstract | Publisher Full Text | Free Full Text

95. Sun YV, Lazarus A, Smith JA, *et al.*: **Gene-specific DNA methylation association with serum levels of C-reactive protein in African Americans.** *PLoS One.* 2013; **8**(8): e73480.
PubMed Abstract | Publisher Full Text | Free Full Text

96. Zhang Y, Yang R, Burwinkel B, *et al.*: **F2RL3 methylation as a biomarker of current and lifetime smoking exposures.** *Environ Health Perspect.* 2014; **122**(2): 131–7.
PubMed Abstract | Publisher Full Text | Free Full Text

97. Panni T, Mehta AJ, Schwartz JD, *et al.*: **Genome-Wide Analysis of DNA Methylation and Fine Particulate Matter Air Pollution in Three Study Populations: KORA F3, KORA F4, and the Normative Aging Study.** *Environ Health Perspect.* 2016; **124**(7): 983–90.
PubMed Abstract | Publisher Full Text | Free Full Text

98. Dai L, Mehta A, Mordukhovich I, *et al.*: **Differential DNA methylation and $PM_{2.5}$ species in a 450K epigenome-wide association study.** *Epigenetics.* 2017; **12**(2): 139–48.
PubMed Abstract | Publisher Full Text | Free Full Text

99. Lindsey ML, Mayr M, Gomes AV, *et al.*: **Transformative Impact of Proteomics on Cardiovascular Health and Disease: A Scientific Statement From the American Heart Association.** *Circulation.* 2015; **132**(9): 852–72.
PubMed Abstract | Publisher Full Text

100. Doll S, Dressen M, Geyer PE, *et al.*: **Region and cell-type resolved quantitative proteomic map of the human heart.** *Nat Commun.* 2017; **8**(1): 1469.
PubMed Abstract | Publisher Full Text | Free Full Text

101. Bom MJ, Levin E, Driessen RS, *et al.*: **Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease.** *EBioMedicine.* 2019; **39**: 109–17.
PubMed Abstract | Publisher Full Text | Free Full Text

102. Hoogeveen RM, Pereira JPB, Nurmohamed NS, *et al.*: **Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention.** *Eur Heart J.* 2020; **41**(41): 3998–4007.
PubMed Abstract | Publisher Full Text | Free Full Text

103. Steenman M, Lamirault G, Le Meur N, Leger JJ: **Gene expression profiling in human cardiovascular disease.** *Clin Chem Lab Med.* 2005; **43**(7): 696–701.
PubMed Abstract | Publisher Full Text

104. Seo D, Wang T, Dressman H, *et al.*: **Gene expression phenotypes of atherosclerosis.** *Arterioscler Thromb Vasc Biol.* 2004; **24**: 1922–7.
PubMed Abstract | Publisher Full Text

105. Robinson EL, Baker AH, Brittan M, *et al.*: **Dissecting the transcriptome in cardiovascular disease.** *Cardiovasc Res.* 2021; cvab117.
PubMed Abstract | Publisher Full Text

106. Cheng ML, Wang CH, Shiao MS, *et al.*: **Metabolic disturbances identified in plasma are associated with outcomes in patients with heart failure: diagnostic and prognostic value of metabolomics.** *J Am Coll Cardiol.* 2015; **65**(15): 1509–20.
PubMed Abstract | Publisher Full Text

107. Zhao J, Zhu Y, Hyun N, *et al.*: **Novel metabolic markers for the risk of diabetes development in American Indians.** *Diabetes Care.* 2015; **38**(2): 220–7.
PubMed Abstract | Publisher Full Text | Free Full Text

108. Newgard CB: **Interplay between lipids and branched-chain amino acids in development of insulin resistance.** *Cell Metab.* 2012; **15**(5): 606–14.
PubMed Abstract | Publisher Full Text | Free Full Text

109. Jones DP: **Sequencing the exposome: A call to action.** *Toxicol Rep.* 2016; **3**: 29–45.
PubMed Abstract | Publisher Full Text | Free Full Text

110. Wang Z, Zhu C, Nambi V, *et al.*: **Metabolomic Pattern Predicts Incident Coronary Heart Disease.** *Arterioscler Thromb Vasc Biol.* 2019; **39**(7): 1475–82.
PubMed Abstract | Publisher Full Text | Free Full Text

111. Bhupathiraju SN, Guasch-Ferre M, Gadgil MD, *et al.*: **Dietary Patterns among Asian Indians Living in the United States Have Distinct Metabolomic Profiles That Are Associated with Cardiometabolic Risk.** *J Nutr.* 2018; **148**(7): 1150–9.
PubMed Abstract | Publisher Full Text | Free Full Text

112. Zheng Y, Yu B, Alexander D, *et al.*: **Associations between metabolomic compounds and incident heart failure among African Americans: the ARIC Study.** *Am J Epidemiol.* 2013; **178**(4): 534–42.
PubMed Abstract | Publisher Full Text | Free Full Text

113. Lanfear DE, Gibbs JJ, Li J, *et al.*: **Targeted Metabolomic Profiling of Plasma and Survival in Heart Failure Patients.** *JACC Heart Fail.* 2017; **5**(11): 823–32.
PubMed Abstract | Publisher Full Text | Free Full Text

114. Bjornsson HT, Fallin MD, Feinberg AP: **An integrated epigenetic and genetic approach to common human disease.** *Trends Genet.* 2004; **20**(8): 350–8.
PubMed Abstract | Publisher Full Text

115. Shin SY, Fauman EB, Petersen AK, *et al.*: **An atlas of genetic influences on human blood metabolites.** *Nat Genet.* 2014; **46**(6): 543–50.
PubMed Abstract | Publisher Full Text | Free Full Text

116. Sun YV, Hu YJ: **Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases.** *Adv Genet.* 2016; **93**: 147–90.
PubMed Abstract | Publisher Full Text | Free Full Text

117. Chen R, Mias GI, Li-Pook-Than J, *et al.*: **Personal omics profiling reveals dynamic molecular and medical phenotypes.** *Cell.* 2012; **148**(6): 1293–307.
PubMed Abstract | Publisher Full Text | Free Full Text

118. Schussler-Fiorenza Rose SM, Contrepois K, Moneghetti KJ, *et al.*: **A longitudinal big data approach for precision health.** *Nat Med.* 2019; **25**(5): 792–804.
PubMed Abstract | Publisher Full Text | Free Full Text

119. Yao C, Chen BH, Joehanes R, *et al.*: **Integromic analysis of genetic variation and gene expression identifies networks for cardiovascular disease phenotypes.** *Circulation.* 2015; **131**(6): 536–49.
PubMed Abstract | Publisher Full Text | Free Full Text

120. Andersson C, Lin H, Liu C, *et al.*: **Integrated Multiomics Approach to Identify Genetic Underpinnings of Heart Failure and Its Echocardiographic Precursors: Framingham Heart Study.** *Circ Genom Precis Med.* 2019; **12**(12): e002489.
PubMed Abstract | Publisher Full Text

121. Palou-Marquez G, Subirana I, Nonell L, *et al.*: **DNA methylation and gene expression integration in cardiovascular disease.** *Clin Epigenetics.* 2021; **13**(1): 75.
PubMed Abstract | Publisher Full Text | Free Full Text

122. Huang Y, Hui Q, Walker DI, *et al.*: **Untargeted metabolomics reveals multiple metabolites influencing smoking-related DNA methylation.** *Epigenomics.* 2018; **10**(4): 379–93.
PubMed Abstract | Publisher Full Text | Free Full Text

123. Hellmuth C, Kirchberg FF, Lass N, *et al.*: **Tyrosine Is Associated with Insulin Resistance in Longitudinal Metabolomic Profiling of Obese Children.** *J Diabetes Res.* 2016; **2016**: 2108909.
PubMed Abstract | Publisher Full Text | Free Full Text

124. Hoffman JM, Tran V, Wachtman LM, *et al.*: **A longitudinal analysis of the effects of age on the blood plasma metabolome in the common marmoset, *Callithrix jacchus*.** *Exp Gerontol.* 2016; **76**: 17–24.
PubMed Abstract | Publisher Full Text | Free Full Text

125. Buchwald P, Tamayo-Garcia A, Ramamoorthy S, *et al.*: **Comprehensive Metabolomics Study To Assess Longitudinal Biochemical Changes and Potential Early Biomarkers in Nonobese Diabetic Mice That Progress to Diabetes.** *J Proteome Res.* 2017; **16**(10): 3873–90.
PubMed Abstract | Publisher Full Text

126. Burt SA, McGue M, Iacono WG: **Nonshared environmental mediation of the association between deviant peer affiliation and adolescent externalizing behaviors over time: results from a cross-lagged monozygotic twin differences design.** *Dev Psychol.* 2009; **45**(6): 1752–60.
PubMed Abstract | Publisher Full Text | Free Full Text

127. Purcell S, Neale B, Todd-Brown K, *et al.*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet.* 2007; **81**(3): 559–75.
PubMed Abstract | Publisher Full Text | Free Full Text

128. Zhan X, Hu Y, Li B, *et al.*: **RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data.** *Bioinformatics.* 2016; **32**(9): 1423–6.
PubMed Abstract | Publisher Full Text | Free Full Text

129. Gogarten SM, Sofer T, Chen H, *et al.*: **Genetic association testing using the GENESIS R/Bioconductor package.** *Bioinformatics.* 2019; **35**(24): 5346–8.
PubMed Abstract | Publisher Full Text | Free Full Text

130. Rohart F, Gautier B, Singh A, *et al.*: **mixOmics: An R package for 'omics feature selection and multiple data integration.** *PLoS Comput Biol.* 2017; **13**(11): e1005752.
PubMed Abstract | Publisher Full Text | Free Full Text

131. Li S, Park Y, Duraisingham S, *et al.*: **Predicting network activity from high throughput metabolomics.** *PLoS Comput Biol.* 2013; **9**(7): e1003123.
PubMed Abstract | Publisher Full Text | Free Full Text

132. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation**

network analysis. *BMC Bioinformatics.* 2008; **9**: 559.
PubMed Abstract | Publisher Full Text | Free Full Text

133. Uppal K, Ma C, Go YM, *et al.*: **xMWAS: a data-driven integration and differential network analysis tool.** *Bioinformatics.* 2018; **34**(4): 701–2.
PubMed Abstract | Publisher Full Text | Free Full Text

134. Redondo MJ, Hagopian WA, Oram R, *et al.*: **The clinical consequences of heterogeneity within and between different diabetes types.** *Diabetologia.* 2020; **63**(10): 2040–8.
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.* 2013; **54**(7): 480–99.
PubMed Abstract | Publisher Full Text

136. Miller GW, Jones DP: **The nature of nurture: refining the definition of the exposome.** *Toxicol Sci.* 2014; **137**(1): 1–2.
PubMed Abstract | Publisher Full Text | Free Full Text

137. Nair M, Ali MK, Ajay VS, *et al.*: **CARRS Surveillance study: design and methods to assess burdens from multiple perspectives.** *BMC Public Health.* 2012; **12**: 701.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 03 June 2021

https://doi.org/10.21956/wellcomeopenres.18643.r43985

✔ **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.

Reviewer Report 24 May 2021

https://doi.org/10.21956/wellcomeopenres.18643.r43986

✔ **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

?

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[1].) 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

1. Wang M, Menon R, Mishra S, Patel AP, et al.: Validation of a Genome-Wide Polygenic Score for Coronary Artery Disease in South Asians.*J Am Coll Cardiol*. 2020; **76** (6): 703-714 PubMed Abstract | Publisher Full Text

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 understating 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 predicter 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 revers 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.

Reviewer Report 13 November 2020

https://doi.org/10.21956/wellcomeopenres.17960.r41143

?

Chunyu Liu
Department of Biostatistics, Boston University, Boston, MA, USA

Cardiovascular disease (CVD) is the leading cause of mortality in South Asia, with rapidly increasing prevalence of hypertension, type 2 diabetes and hyperlipidemia over the last two decades. A thorough understanding of CVD-associated omics profiles may lead to more accurate estimates of CVD progression and stimulate new strategies for improving cardiovascular health in South Asia. However, I don't think this review provided a comprehensive review for multi-omics study of CVD and risk factors. Below is the summary of my comments.

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.

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.

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-ethic 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.